Energy Research and Development Division
FINAL PROJECT REPORT

A SEASONAL PERSPECTIVE ON REGIONAL AIR QUALITY IN CENTRAL CALIFORNIA AND IMPACTS OF ELECTRICITY GENERATION

Prepared for: California Energy Commission
Prepared by: Lawrence Berkeley National Laboratory

DECEMBER 2011
CEC-500-2013-035

PREPARED BY:

Primary Author(s):
Ling Jin
Nancy J. Brown

Lawrence Berkeley National Laboratory
Berkeley, CA 96270

Contract Number: 500-02-004

Prepared for:

California Energy Commission

Marla Mueller
Contract Manager

Guido Franco
Program Area Lead
Energy-Related Environmental Research

Linda Spiegel
Office Manager
Energy Generation Research Office

Laurie ten Hope
Deputy Director
Energy Research and Development Division

Robert P. Oglesby
Executive Director

ACKNOWLEDGEMENTS

The authors wish to thank the following people for their help and advice:

- Robert A. Harley (University of California, Berkeley)
- Jim Wilczak, Jian-Wen Bao, and Sara Michelson (National Oceanic and Atmospheric Administration)
- Allison Steiner (University of California, Berkeley)
- Neva Lowery, Bruce Jackson, Ajith Kaduwela, Jin-You Liang, Klaus Scott, Patricia Velasco, and Eugene Yang (California Air Resources Board)
- Daniel Cohan (Georgia Environmental Protection Division)
- Philip Martien (Bay Area Air Quality Management District)
- Stephen Ziman (Chevron)
- Steven Reynolds (Envair)
- Shaheen Tonse, Xiaoling Mao, Claire Agnoux, Marie-Anne Serve, and Sarah M. Sermondadaz (Lawrence Berkeley National Laboratory)

PREFACE

The California Energy Commission Energy Research and Development Division supports public interest energy research and development that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy services and products to the marketplace.

The Energy Research and Development Division conducts public interest research, development, and demonstration (RD&D) projects to benefit California.

The Energy Research and Development Division strives to conduct the most promising public interest energy research by partnering with RD&D entities, including individuals, businesses, utilities, and public or private research institutions.

Energy Research and Development Division funding efforts are focused on the following RD&D program areas:

- Buildings End-Use Energy Efficiency
- Energy Innovations Small Grants
- Energy-Related Environmental Research
- Energy Systems Integration
- Environmentally Preferred Advanced Generation
- Industrial/Agricultural/Water End-Use Energy Efficiency
- Renewable Energy Technologies
- Transportation

A Seasonal Perspective on Regional Air Quality in Central California and Impacts of Electricity Generation is the final report for the Seasonal Modeling project (Contract Number 500-02-004, Work Authorization Number MR-026, WA MAQ-05-02) conducted by Lawrence Berkeley National Laboratory. The information from this project contributes to Energy Research and Development Division's Energy- Related Environmental Research Program.

For more information about the Energy Research and Development Division, please visit the Energy Commission's website at www.energy.ca.gov/research/ or contact the Energy Commission at 916-327-1551.

ABSTRACT

Population within the Central California region is growing rapidly, and demand for electricity and new power generation sources are increasing. At the same time, there are persistent, serious air pollution problems, including fine particulate matter and ozone. Improved analysis and prediction of air quality impacts from emissions sources such as power plants are required for planners and regulators to more accurately determine the true impact of various emissions sources on air quality, and to evaluate alternative control strategies such as short-term curtailment of emissions. This research sought a comprehensive evaluation of air quality model performance and development of analysis methods to characterize air quality and ozone, including ozone from electrical generation, on a new level. This study's results indicate that high ozone occurs during multiple types of summer meteorological flow patterns, underscoring the need for an approach that is more inclusive than episodic modeling. Additional methods applied in this project, including cluster characterization and sensitivity analysis, are powerful techniques for understanding the factors responsible for high ozone levels. By applying these methods, a half dozen high ozone meteorological regimes in the San Joaquin Valley were distinguished, and these can be used to evaluate on a broader level the impacts of potential and even unconventional control strategies for attaining air quality standards. For example, this approach has the potential to illustrate and quantify that the location and timing of power production can be just as important as the plant's size and emission rate. While this factor has always been conceptually known, conventional tools cannot measure this effect over the long term, and hence it cannot currently be considered a quantifiable method of controlling pollution.

Keywords: Ozone, air quality, air pollution, emissions, model, Central California

Please use the following citation for this report:

Jin, Ling, and Nancy J. Brow (Lawrence Berkeley National Laboratory). 2011. ***A Seasonal Perspective on Regional Air Quality in Central California and Impacts of Electricity Generations.*** California Energy Commission. Publication Number: CEC-500-2013-035

TABLE OF CONTENTS

Acknowledgements ..... **i**

PREFACE ..... **ii**

ABSTRACT ..... **iii**

TABLE OF CONTENTS ..... **iv**

LIST OF FIGURES ..... **vi**

LIST OF TABLES ..... **ix**

EXECUTIVE SUMMARY **1**

Introduction 1

Purpose 1

Model Performance Evaluation 2

Cluster Analysis: Meteorological Regimes and Associated Spatial Patterns for Ozone 2

Sensitivity Analysis 3

Conclusions 3

Recommendations for Future Research 4

Benefits to California 5

CHAPTER 1: Introduction **7**

1.1 Context and Motivation 7

1.2 Research Scope and Objectives 8

1.3 Approach 9

CHAPTER 2: Performance Evaluation of MM5-CMAQ for the SARMAP Domain **11**

2.1 Introduction 11

2.2 Model Description 11

2.2.1 Modeling Domain 11

2.2.2 Meteorological Inputs 13

2.2.3 Emission Inputs 13

2.2.4 Community Multi-scale Air Quality Model 17

2.2.5 Model Evaluation and Diagnostic Methods 18

2.3 Initial Simulations and Model Improvements 20
2.3.1 Initial Simulations and Diagnostics 20
2.3.3 Refining Input Parameters 24
2.4 Evaluation of Base Case Simulations 26
2.4.1 Spatial and Temporal Trends in Model Performance 26
2.4.2 Effects of Meteorology Nudging on Air Quality Simulations 29
2.4.3 Ozone Production Efficiency 30
2.5 Final Sensitivity Analysis and Discussions 33
CHAPTER 3: Sensitivity Analysis of Ozone Formation and Transport 35
3.1 Introduction 35
3.2 Methods 36
3.2.1 Study Domain and Input Data 36
3.2.2 CMAQ-HDDM Modeling System 37
3.2.3 Ozone Control Regimes 38
3.3 Results and Discussions 39
3.3.1 Simulated Ozone and First-order Sensitivity Analysis 39
3.3.2 Ozone Control Regimes in the SJV 41
3.3.3 Local Versus Upwind Source Contributions 43
3.3.4 Influence of Uncertainties in Other Parameters 48
CHAPTER 4: Seasonal Versus Episodic Performance Evaluation for the CCOS Domain 52
4.1 Introduction 52
4.2 Data and Methods 53
4.2.1 Study Domain 53
4.2.2. Emission Inputs 53
4.1.1 Meteorological Fields 61
4.1.2 Intensive Operating Periods (IOPs) 61
4.2 Air Quality Model Simulations and Performance Evaluation 62
4.2.1 Simulated Ozone Concentrations 62

4.2.2 Temporal and Spatial Behaviors of Ozone Predictions 63
4.2.3 Model Performance for Different Synoptic Events 69
4.3 Conclusions 82
CHAPTER 5: Clustering Regional Ozone Concentrations to Reveal Meteorological Regimes Influencing Air Quality 84
5.1 Introduction 84
5.2 Methods 86
5.2.1 Principal Component Analysis 86
5.2.2 Cluster Analysis Methods 88
5.2.3 Data and Subregions 89
5.3 Results 91
5.3.1 Ozone Regimes in Central California 91
5.3.2 Ozone Regimes in the SJV 95
5.3.3 Ozone Regimes in the SFB and the SV 98
5.3.4 Meteorological Features by Ozone Regimes 104
5.4 Discussion 113
5.4.1 Effects of Weekend Emission Reductions on Spatial Distribution of Ozone 113
5.4.2 Representativeness of Ozone Observational Sites 114
5.4.3 Comparison with Historical Data 116
CHAPTER 6: Conclusions and Recommendations 118
6.1 Summary of Major Findings 118
6.2 Recommendations for Future Research 119
GLOSSARY 122
REFERENCES 125
APPENDIX A: Quality Assurance of CCOS Emissions A-1
APPENDIX B: Supporting Materials for Ozone Cluster Analysis B-1

LIST OF FIGURES

Figure 2-1: Modeling Domain Indicated by the Purple Rectangle 12

Figure 2-2: Description of the Outer (36-km), Middle (12-km) and Inner (4-km) Nested Grids Used in the MM5 Simulation 13

Figure 2-3: NOx Emission Profiles Summed over the Domain for the Various Categories 15

Figure 2-4: CO Emission Profiles Summed over the Domain for the Various Categories 16

Figure 2-5: VOC Emission Profiles Summed over the Domain for the Various Categories 16

Figure 2-6: Spatial Distribution of NOx Emissions at Noon PDT 17

Figure 2-7: Taylor Diagram Example 19

Figure 2-8: Kriged Daytime Ozone Normalized Biases (%) 21

Figure 2-9: Ozone Biases (ppb) Histograms (Day and Night) 21

Figure 2-10: Ozone and Odd Oxygen (O_3 + NO_2) Time Series Plots 23

Figure 2-11: Taylor Diagram Showing Daytime Pattern Statistics 27

Figure 2-12: Daytime Normalized Biases at Three Air Basins 28

Figure 2-13: Spatial Distribution of Ozone Biases (%) at 3 p.m. on Day 214 30

Figure 2-14: Ozone Production Efficiency (the Linearly Fitted Slope) along with Standard Errors (se) 32

Figure 2-15: Mean Absolute Fully Normalized Sensitivity Coefficients of Ozone 33

Figure 3-1 Modeling Domain Indicated by the Purple Rectangle 36

Figure 3-2: Weekday Average Peak Simulation Results (in ppb) 40

Figure 3-3: Ozone Control Options for Peak Ozone Exceedances During the Five-Day Episode 41

Figure 3-4: Mapping of Ozone Sensitivity Options at 8-h Peak Ozone Exceedances in the SJV .. 43

Figure 3-5: Weekday Average 8-h Peak Ozone Sensitivities and Decomposition of Contributions from Subregions 45

Figure 3-6: Contributions to $\frac{\partial[O_3]}{\partial \varepsilon_{E_{SFBAVOC}}}$ and $\frac{\partial[VOC]}{\partial \varepsilon_{E_{SFBAVOC}}}$ 47

Figure 3-7: Weekday Average Second-order Sensitivities of 8-h Peak Ozone 49

Figure 3-8: Weekday Average of Second-Order Sensitivity Coefficients of 8-h Peak Ozone Along East and West Transects 51

Figure 4-1: SARMAP and CCOS Domain 55

Figure 4-2: Daily Total Emissions in Summer 2000 56

Figure 4-3: Daily Total Isoprene and Terpene Emissions in Summer 2000 57

Figure 4-4: Location of California Forest Fires During Summer 2000 58
Figure 4-5: Daily Total Forest Fire Emissions for the CCOS Domain During Summer 2000 60
Figure 4-6: Simulated 8-h Peak Ozone (ppb) 63
Figure 4-7: Ozone and Evaluation Metrics by Day 64
Figure 4-8: Distribution of 1-h Ozone Peak Time 67
Figure 4-9: Comparison of Modeled and Observed 8-h Peak Ozone 68
Figure 4-10: Location of Wind Profiler Clusters Used for MM5 Model Evaluation 70
Figure 4-11: Time-Height Series Plot for Meteorological Evaluation: Summer 71
Figure 4-12: Time-Height Series Plot for Meteorological Evaluation: IOP 1 72
Figure 4-13: Time-Height Series Plot for Meteorological Evaluation: IOP 2 73
Figure 4-14: Time-Height Series Plot for Meteorological Evaluation: IOP 3 74
Figure 4-15: Time-Height Series Plot for Meteorological Evaluation: IOP 4 75
Figure 4-16: Pattern Statistics for IOPs and the Whole Season 77
Figure 4-17: Normalized Bias Surface of 8-h Peak Ozone Predictions 80
Figure 5-1: Three Subregions in Central California for Clustering Ozone Spatial Patterns 90
Figure 5-2: Mean 8-h Ozone Maxima (ppb) Maps for the Three Subregions 92
Figure 5-3: Ozone Loadings (ppb) on the Three PCs of 8-h Ozone Maxima 93
Figure 5-4: Visualizing SJV Ozone Clusters (1 to 6) with PC Scores 96
Figure 5-5: Ozone Anomaly Field (ppb) Patterns Averaged over SJV Clusters 97
Figure 5-6: Mean Ozone Anomaly Fields (ppb) Patterns of SFB Clusters 99
Figure 5-7: Mean Ozone Anomaly Fields (ppb) Patterns of SV Clusters 100
Figure 5-8: Cluster Memberships 103
Figure 5-9: Average Temperature Anomalies (K) for Each SJV Ozone Cluster 106
Figure 5-10 Average Wind Fields and First Two PC Loadings 108
Figure 5-11: Diurnal Profiles of Surface Wind Scores 109
Figure 5-12: SJV Weekend Effects on Ozone Spatial Patterns Projected onto the First Three PCs 114
Figure A1: California Air Basins A-1
Figure A2: CCOS Emission Inputs Compared to ARB Inventory: Mobile Sources A-4

Figure A3: CCOS Emission Inputs Compared to ARB Inventory: Stationary Sources (Area+Point) A-4

Figure A4: CCOS Emission Inputs Compared to ARB Inventory: Biogenic Sources (Tons/Day) .. 5

Figure A5: Abnormal Temporal Profile in ALK4 from Mobile Sources A-5

Figure A6: Locations Where Abnormal Temporal Patterns Occur in ALK4 Emissions A-6

Figure A7: Distribution of Time and Date When the Abnormal Pattern Occurs A-6

Figure B1: Flow Diagram Describing the Clustering Procedure B-2

Figure B2: Variance and Cumulative Variance Explained by Principal Components of 8-h Ozone Maxima in Each Subregion: (a) SJV, (b) SFB, (c) SV B-3

Figure B2: Continued B-4

Figure B3: Scatter Plots of Ozone Days on PC1 and PC2. Cluster indices are shown in the same color as the associated days at the respective cluster centers. B-5

Figure B4: Average Daily Maximum Temperature (K) for Summer 2000 and the Temperature Loadings (K) on the First Three PCs B-6

Figure B5: Variances and Cumulative Variances Contributed by Wind PCs B-7

Figure B6: Mean Sea Surface Pressures and Anomolies (at Noon) Averaged Over Each SJV Cluster B-7

LIST OF TABLES

Table 2-1 Evaluation and Diagnostic Tools 18

Table 2-2: Summary of Initial Model Diagnostic Simulations 22

Table 2-3: Boundary Conditions for Selected SAPRC-99 Species 24

Table 2-4: Peak Ozone (1-h and 8-h) Predictions Under Old and Improved (Base Case) Settings 25

Table 2-5: Summary Statistics for Three 5-Day Periods 29

Table 2-6: Normalized Bias and Gross Error for Model Simulations 30

Table 3-1 Emission Inputs from Selected Regions 37

Table 4-1: Performance Statistics. 66

Table 4-2: Summary of Model Performance 78

Table 5-1: Description of the First Three PCs for Each Subregion 94

Table 5-2: Descriptions of Ozone Clusters 101

Table 5-3: Inter-Basin Relationships of Ozone Clusters 104

Table 5-4: Characteristics of SJV Moderate to High Ozone Clusters 112

Table 5-4: Comparison of SJV Clustering Results with Constant Emissions and Day-of-Week Emissions 114

Table 5-5: Comparison of Clustering Results with Modeled Data at All the Grids vs. Using Modeled Data Only at Observational Sites 115

Table A1: Mapping of Chemical Species A-2

Table A2: Carbon Numbers for VOC Species A-3

Table B1: Transition Matrix for SJV Ozone Clusters B-5

EXECUTIVE SUMMARY

Introduction

Central California spans a wide variety of urban, agricultural, and natural terrain, including the San Francisco Bay area, the Central Valley, and the Sierra Nevada Mountains. Population within this region is growing rapidly, and demand for electricity and new power generation sources are increasing. At the same time, there are persistent, serious air pollution problems, including fine particulate matter ($PM_{2.5}$) and ozone. Reducing ozone concentrations has been especially challenging for the San Joaquin Valley. In addition to its diverse local emissions, the San Joaquin Valley is affected by pollutants emitted or formed in upwind source regions, which further complicates understanding the source-receptor relationships important for identifying potential control strategies.

Current practice to control ozone relies on simulating a few ozone episodes (three to five day periods) with "worst case" weather conditions. More important, due to the diverse meteorological conditions in Central California, it is not known whether control strategies developed from the worst-case situations will be effective for different meteorology and emissions. Air quality analyses based on modeling an entire summer season or longer are no longer limited because of lack of adequate supporting data for model input preparation and evaluation, and high computational costs. Furthermore, analyzing outputs for an entire summer season, rather than more focused ozone episodes, is challenging.

Purpose

This research sought a comprehensive evaluation of air quality model performance and development of analysis methods to characterize air quality and ozone, including ozone from electrical generation, on a new level. For example, through statistical grouping of important weather patterns known as "clustering," this approach can provide a more comprehensive view than single episodic analysis in guiding control strategy design for State Implementation Plans, within acceptable processing timelines. The analysis approach is adept at quickly distinguishing the important role of location and timing of emissions; namely, that *x* tons of nitrogen oxides (NOx) at time *A* may *increase* ozone, but the same amount of NOx released at time *B* may *reduce* ozone. This type of tool could eventually be used to assess potential unconventional control strategies such as strategic placement of power plants and other sources, or emissions curtailment for certain periods.

This research is focused on the Central Valley. However, the methods developed in this project, including the model performance evaluation, cluster characterization, and sensitivity analysis, have wide applications for other modeling studies using different pollutants and locations. Although this approach has been shown to be promising in terms of acceptable performance characteristics and processing time, there are several areas of model development and input data improvements that may be necessary before the process could be considered for use in planning. These areas are discussed in the Recommendations section of this Executive Summary.

Model Performance Evaluation

This study evaluated in detail the seasonal and episodic performance of the Community Multiscale Air Quality Model applied to simulate air quality at a fine-grid spacing (4 kilometer horizontal resolution) in Central California, where ozone air pollution problems are severe. A rich aerometric database collected during the summer 2000 Central California Ozone Study was used to prepare model inputs and to evaluate meteorological simulations and chemical outputs.

An important result of this study is a suite of benchmarked model inputs that describe emissions and meteorological variations in both time and space for Central California for an entire summer season. A variety of evaluation methods and diagnostic tools have been applied to refine model inputs and to evaluate model performance. Initially this was done for a 15-day period, and then the modeling was extended to the entire summer. Lateral boundary conditions, minimum vertical diffusivity, and ozone dry deposition rate over the ocean were refined according to model diagnostics, observations, and published literature. Gridded meteorological and emission inputs were developed to reflect variability occurring on diurnal, weekly, and seasonal time scales. Driven by these inputs, the model demonstrated stable performance for the entire summer without further adjustment of input parameters. High-ozone events are highlighted (exemplified by four intensive operating periods of the Central California Ozone Study) for evaluating both meteorological inputs and chemical outputs (ozone and its precursors) and comparing them to the summer average.

For most of the summer days, normalized gross errors in the modeling domain were found to be less than 25 percent for modeled hourly ozone, and normalized biases were within ±15 percent for both ozone maxima (1-hour and 8-hour). The domain-wide grouped metrics indicated similar performance between the intensive operating periods and the whole summer, with respect to predicted ozone and its precursors. Episode-to-episode differences in ozone predictions were more pronounced at a subregional level. The model performed consistently better when applied to the San Joaquin Valley than for the other air basins, and for episodic ozone predictions that were similar to the summer average. The model performed less well, with normalized peak ozone biases outside the range of ±15 percent, when applied to the Sacramento Valley and the San Francisco Bay area. Problems were most noticeable in episodes that were subject to the largest uncertainties in meteorological fields, such as wind directions in the Sacramento Valley, and timing and strength of onshore flow in the San Francisco Bay area.

Cluster Analysis: Meteorological Regimes and Associated Spatial Patterns for Ozone

Distinctive meteorological regimes that are associated with different ozone spatial patterns in the San Joaquin Valley were determined using cluster analysis and principal component analysis. *Cluster analysis* is a statistical method that groups large amounts of data into smaller categories, such that each subset shares a common trait. The common traits used in the cluster analysis for this study were meteorological effects shown to explain the observed patterns of ozone concentration within the modeling domain, their dynamic relationship with each other, and their relationship to those in upwind areas. In general, average ozone levels seen in the San Joaquin Valley increase with temperature, while their spatial distributions depend on flow regimes, especially the strength of sea breezes and upslope flows. A high-ozone episode in late

July/early August was shown to have a different flow regime that was distinct from other days in summer 2000, when significant ozone exceedances occurred under more stagnant conditions. Ozone spatial patterns and the associated weather regimes captured in these simulations were reflected in those derived from a historical observational data set, which provided an indication of representativeness for the modeling period, and indicated that the model reproduced observed ozone patterns correctly. These results can serve as a basis to study variability in ozone responses to emission controls and inter-basin pollutant transport under and across different temperature and flow regimes.

Sensitivity Analysis

A high-order direct decoupled method for conducting sensitivity analysis was demonstrated to be an effective tool for delineating ozone responses to ozone precursor reductions in time and space, and for determining the relative importance of emission contributions from sources such as power plants to ozone concentrations from local versus upwind areas. This study used second-order sensitivity analyses to determine influences of model input uncertainties on ozone responses to emission changes. During the late July/early August ozone episode, NOx emission reductions were overall more effective than reductions of volatile organic compounds for attaining the 8-hour ozone standard in the San Joaquin Valley. In contrast, volatile organic compound reductions would have worked better for attaining the prior 1-hour ozone standard. Inter-basin source contributions of NOx emissions are limited to the northern part of the San Joaquin Valley, while anthropogenic volatile organic compound emissions, especially those emitted at night, influence ozone formation in the San Joaquin Valley further downwind. Among model input parameters studied here, uncertainties in emissions of NOx and anthropogenic volatile organic compounds, and the rate coefficient of the hydroxide + nitrogen dioxide ($OH+NO_2$) termination reaction, are found to have the greatest effects on first-order ozone responses to changes in NOx emissions. Uncertainties in biogenic volatile organic compound emissions only have a modest effect because they are generally not co-located with anthropogenic sources in the San Joaquin Valley.

In a parallel effort, this study's researchers investigated several inter-basin trading scenarios based on simulations conducted for the July-August episode, where emissions were removed from one location in one basin (the subtracting location) and added to a power plant in another basin (the receiving location). This phase of the project is discussed in the report *Impact of Short–Term, Interbasin and Interpollutant Credit Trading on Air Quality and Credit Prices.* From a modeling standpoint, the ozone impacts of trading scenarios proved to be very sensitive to changes in time and space due to the complex chemical reactions, meteorology, and transport in Central California. This exercise underscores the need for additional analysis and improvements in modeling inputs and performance before modeling could be considered a viable method for recommending offset ratios and assessing long-term air quality impacts of potential trades.

Conclusions

- Improved analysis and prediction of air quality impacts from emissions sources such as power plants are required for planners and regulators to more accurately determine the

true impact of various emissions sources on air quality, and to evaluate alternative control strategies such as short-term curtailment of emissions. This research considered some of the limitations of modeling only a few high-ozone episodes as the basis for these types of planning and evaluations. This study's results indicate that high ozone occurs during multiple types of summer meteorological flow patterns, underscoring the need for an approach that is more inclusive than episodic modeling.

- The methods applied in this project, including cluster characterization and sensitivity analysis, are powerful techniques for understanding the factors responsible for high ozone levels. By applying these methods, a half dozen high ozone meteorological regimes in the San Joaquin Valley were distinguished, and these can be used to evaluate on a broader level the impacts of potential and even unconventional control strategies for attaining air quality standards. For example, this approach has the potential to illustrate and quantify that the location and timing of power production can be just as important as the plant's size and emission rate. While this factor has always been conceptually known, conventional tools cannot measure this effect over the long term, and hence it cannot currently be considered a quantifiable method of controlling pollution.
- Variations in ozone spatial distributions in the San Joaquin Valley can be mostly explained by meteorology, with only minor effects from weekend emission changes. Especially sensitive to changes in meteorology include the limiting reagents in the northern San Joaquin Valley and relative impacts of inter-basin transport.
- This study showed that ozone levels along the western side of the San Joaquin Valley can change significantly, depending on the degree of stagnation that is characterized by the strength of sea breeze and ventilation flows to the Sierra Nevada Mountains. This spatial feature has not been captured by the existing ozone monitoring stations, which are located in the more populated middle and eastern side of the San Joaquin Valley. The western part of the San Joaquin Valley is dominated by agricultural land uses, and crop productivity can be adversely affected by ambient ozone levels. Ozone levels in this area are affected both by local mobile sources traveling on Interstate 5 and by pollutants transported from the San Francisco Bay area and other coastal regions.
- The model performance evaluation showed that ozone predictions in the San Francisco Bay area and Sacramento Valley agreed less well with observations during periods that were subject to the largest uncertainties in meteorological fields (such as wind directions in the Sacramento Valley, and timing and strength of onshore flow in the San Francisco Bay area).

Recommendations for Future Research

1. **Expand sensitivity analysis to a broader meteorological range.** Variability in ozone responses to emission changes and inter-basin contributions, and thus emission trading feasibilities, should be quantified by extending the direct decoupled method sensitivity analysis. Since summer 2000 was cooler than usual, ozone production rates in this

period may not be representative of those under more typical summer conditions. Ozone sensitivities should be examined with second-order sensitivity analysis to determine the influence of temperature changes on ozone responses to emission reductions.

2. **Expand regime/cluster analysis to include aloft variables.** Weather patterns associated with different ozone clusters were examined only with surface meteorological variables in this research, which assumes that the prevailing meteorological regimes have significantly affected the ground-level data, from which the weather patterns can be inferred. A more complete description of mesoscale (hundreds of kilometers) and synoptic scale (thousands of kilometers) regimes could be determined by examining upper-level atmospheric conditions. As an example, the evolution and localization of certain high-pressure systems may help to explain the variability within an ozone cluster and its dynamic relationship with the other ozone clusters to which it transitions.

3. **Add air monitors along the Western San Joaquin Valley.** Additional measurements of pollutants along the western part of the San Joaquin Valley could be used to enhance the understanding of crop damage by ozone and its precursors, and to assess changes in ozone concentrations as a result of reducing emissions from local and upwind sources.

4. **Add meteorological measurements in air basins near the San Joaquin Valley.** Wind profilers deployed in Central California during the Central California Ozone Study provided better spatial coverage in the San Joaquin Valley than in other air basins. Better spatial distribution of these measurements is needed to improve the understanding and characterization of wind flow and other boundary layer dynamics in the Sacramento Valley and San Francisco Bay area, and these areas can subsequently benefit from the ozone modeling.

5. **Evaluate uncertainties in the emission inventory, which was not a focus of this research.** Anthropogenic emissions can be affected by meteorological factors. For example, evaporative volatile organic compound emissions increase with ambient temperature. In the current research, the same weekly cycle was used to simulate ozone for the entire summer. The anthropogenic emission inputs could be improved by imposing day-specific variability that is driven by meteorology. Such variability could be important for simulating ozone for an entire summer, during which significant temperature changes occur. Human activities during the summer could also be different from month to month and from day to day, particularly for August, when many people take vacations. The diurnal emission profiles used to allocate total daily emissions to different hours of the day should be investigated to address the unexplained positive bias in ozone predictions at night, after titration effects have been considered.

Benefits to California

This research addresses the Public Interest Energy Research (PIER) goals of improving the environment, public health, and safety by evaluating and recommending improvements to current regulatory methods for evaluating effects of emissions sources such as power plants,

and for possible control measures on air quality and energy efficiency. Specifically, an improved process able to capture spatial and temporal changes in pollution over a whole season would allow for the evaluation of novel control strategies, such as short-term curtailments or trading of emissions, or placement of power plants in less sensitive ozone regimes. This research is one step toward this long-term goal.

CHAPTER 1:
Introduction

1.1 Context and Motivation

Ozone is designated as a criteria pollutant by the Federal Clean Air Act (U.S. EPA 1994) due to its adverse effects on human health (Lippmann 1989, 1993; Bell et al. 2004; Bell and Dominici 2008) and on agriculture productivity (e.g., Morgan et al. 2003). It is also a powerful atmospheric oxidant. Ozone (O_3) and its photolysis also affect climate, because ozone is a greenhouse gas and its photolysis products strongly influence the oxidant content of the atmosphere, which in turn, affect the lifetimes of other important greenhouse gases, such as methane (Stevenson et al. 2000).

Reducing ozone is challenging because it is a secondary pollutant, formed by chemical reactions involving volatile organic compounds (VOCs) and carbon monoxide (CO) in the presence of nitrogen oxides (NOx) and sunlight. Although it is tempting to infer that the severity of ozone pollution could be reduced in a given area by lowering emissions of VOCs, CO, and NOx, or any combination of them, this is not the case, due to the complicated and non-linear nature of its formation chemistry. The rate of ozone production can be limited by either VOCs or NOx. If the term *VOC-limited* is used, it infers that reducing VOCs will result in the reduction of ozone; while, at the same time, a so-called NOx disbenefit occurs, in that ozone concentration actually increases as NOx decreases. In a *NOx-limited* regime, such as forested rural areas with abundant biogenic VOC emissions, ozone concentrations decline with reductions in NOx emissions but show little response to reductions in VOCs. Which of the two precursors is limiting is indicated by formation of peroxides (low NOx) versus nitric acid (high NOx) (Sillman and Wofsy 1995). When the two production rates are comparable, it is in the transition regime where the reduction of either precursor is helpful; however, in this regime, ozone reductions are not as great as they are with only one limiting precursor.

Reducing ozone concentrations is especially challenging for Central California, which has suffered from serious air quality problems due to its unique trough-like geography and diverse emission sources. Over the past decade, there have been over 30 days/year in exceedance of the old 1-hour ozone standard (120 parts per billion, ppb), and even more (>100 days/year) in exceedance of the current 8-hour ozone standard (84 ppb). A more stringent 8-h ozone standard (75 ppb) has been established by the United States Environmental Protection Agency (U.S. EPA) (U.S. EPA 2008), suggesting that more of the region might to be out of compliance with the new standard. The situation is complicated because ozone control regimes of NOx-, and VOC-limitation and transitional behavior change in both time and space as a function of emissions and meteorology (Blanchard and Stoeckenius 2001; Baertsch-Ritter et al. 2004; Couach et al. 2004). The San Joaquin Valley (SJV) also contains pollutants emitted or formed from many upwind source regions, as well as diverse local emissions. This further confounds source-receptor relationships, which are crucial for identifying potential control strategies.

Photochemical air quality models integrate scientific understanding of how pollutants evolve in the atmosphere at regional or larger scales, and have played an important role in developing air quality management plans (Russell and Dennis 2000), such as State Implementation Plans (SIP). Current practice to develop control strategies relies on simulating a few ozone episodes with the "worst case" weather conditions. There are concerns about models being tuned to perform well by adjusting input data and model parameters for specific emission and meteorological conditions. More important, different days may have quite different temperature distributions in space and time, and flow patterns at the surface and aloft, which influence the fate of emissions (e.g., biogenic VOCs), chemistry (e.g., VOC- or NOx-limitation), and transport (e.g., sensitivity to upwind sources). Due to the diverse meteorology in the SJV and other regions in Central California, it is not known whether control strategies developed from the worst-case situations are effective for different meteorologies and emissions. The observational analysis performed by Blanchard and Stoekenius (2001) suggests that ozone control regimes of NOx-, and VOC-limitation, and transitional behavior, change in both time and space. This has not been captured fully in episodic modeling. Hence, modeling for longer term (e.g., a summer season) is of interest to characterize better the variability in relationships of emissions, chemistry, and transport in ozone formation for a range of different weather conditions to derive more integrated strategies for Central California.

Air quality analyses based on modeling an entire summer season or longer periods are limited (e.g., Winner and Cass 2000; Gego et al. 2008), which is mainly due to lack of adequate supporting data for model input preparation and evaluation, as well as high computational cost. Furthermore, analyzing outputs for a period of an entire summer season, rather than a more focused ozone episode, is challenging. Previous studies demonstrated several analysis approaches for seasonal or annual model simulations, but focused on temporal features (e.g., Biswas et al. 2001; Hogrefe et al. 2006). A structured data analysis framework is needed to extract efficiently information associated with various types of ozone behaviors captured in the summer season, both in time and space.

1.2 Research Scope and Objectives

This research seeks a comprehensive evaluation of air quality model performance and characterization of ozone variability through application of the Community Multiscale Air Quality Model (CMAQ) to Central California for an entire summer season. Summer 2000 was chosen for this study because the rich aerometric database collected during this time period for the Central California Ozone Study (CCOS) can be used for model input preparation, model evaluation, and corroborative analysis of the model results. To date, modeling efforts have been focused on a few ozone episodes. When used for preparing State Implementation Plans, model simulations of these episodes underlie our knowledge of issues such as ozone responses to emission-reduction strategies and inter-basin versus intra-basin contributions to ozone concentrations, which are important for developing effective control strategies. This study indicates limitations in modeling only worst-case scenarios by demonstrating the differences in pollutant production, transportation, and accumulation regimes that are characterized with distinctive spatial patterns of high ozone days and their associated emission and meteorological

conditions across the entire summer season. The findings indicate that, when we use air quality modeling to guide control strategy design, it is beneficial to examine patterns that occur in space and time rather than what happens at a single point in time.

The first specific objective is to determine the CMAQ configuration (e.g., chemical solvers, advection scheme) and model inputs (e.g., nudged wind versus unnudged wind, constant versus vertically varying boundary conditions) that are appropriate for the study domain. The second objective is to demonstrate the suitability of an analysis approach with the High-order Decoupled Direct Method (HDDM) within the CMAQ model to investigate ozone responses to precursor emissions from both local and upwind air basins. Then, we extend the modeling period to the entire summer season, to evaluate seasonal and episodic model performances under various emission and meteorological conditions and to characterize ozone patterns and associated weather regimes with cluster analysis.

This study represents the first attempt to simulate air quality at a fine spatial resolution for time scales longer than ozone episodes. The new methods aim to understand impacts of emissions sources (including electrical generation) on air quality on a whole new level. Improved methods would allow planners and regulators to more accurately determine the true impact of various sources on air quality, as well as to evaluate alternative control strategies such as short-term curtailment of emissions. Although the study derives results for specific regions and time periods, the methods developed in this project, including the model performance evaluation, cluster characterization, and sensitivity analysis, have wide applications for other modeling studies using different pollutants and locations.

1.3 Approach

Chapter 2 of this report presents a diagnostic and mechanistic evaluation of the Mesoscale Modeling System (MM5)-CMAQ model for the San Joaquin Valley Air Quality Study (SJVAQS)/Atmospheric Utility Signatures, Predictions and Experiments (AUSPEX) Regional Modeling Adaptation Project (SARMAP) domain. That domain covers the entire SJV, San Francisco Bay area, and the metropolitan area of Sacramento (part of the Sacramento Valley), as well as other air basins. Although it is a subset of the larger CCOS domain, the SARMAP domain has been used in past modeling studies (e.g., Steiner et al. 2006). Sensitivity analyses are performed with the brute force method and Direct Decoupled method (DDM) to diagnose causes for differences between observations and model predictions. Model inputs and parameters are refined and evaluated based on model diagnostics, observations, and the literature.

Chapter 3 contains a description of an ozone sensitivity analysis method with application of HDDM-CMAQ to the SARMAP domain. We determine changes in limiting reagent (NOx or VOCs) with time and space, importance of local versus upwind emission sources, and the effects of emission timing. We assess the influences of uncertainties in emission inventories, reaction rate coefficients, and boundary conditions on the effects of potential anthropogenic emission-reduction strategies with second-order sensitivity analyses.

Chapter 4 describes the extension of the model. The modeling period is extended to the entire summer season that lasts for 117 days and the modeling domain is extended to the entire CCOS domain, which covers all of Central California and most of Northern California. Emission inputs are processed for the CCOS domain, and a comprehensive evaluation of model performance for predicting 1-h peak, 8-h average ozone concentrations, and ozone precursors is performed across a range of days and locations, with significant underlying spatial and temporal variations in air quality for the entire summer.

Chapter 5 describes a characterization of ozone spatial patterns and associated meteorological regimes for the modeling period obtained from cluster analysis and principal component analysis. Distinctive ozone patterns are determined for the San Joaquin Valley, as well as the Sacramento Valley and the San Francisco Bay area, based upon the spatial distributions of daily 8-h ozone maxima. We assess the dynamic relationship among ozone patterns in the SJV and their relationship to those in upwind areas. Temperature and flow patterns that are associated with each ozone cluster are determined, and their joint effects were shown to explain the observed spatial patterns in each ozone cluster. The worst-case ozone episode is shown to have a rather different flow regime that is distinct from some other days when significant ozone exceedances occur. Ozone patterns and the associated weather regimes are compared with those derived from a historical observational data set, to ensure adequate model performance and assess representativeness of the modeling season.

The concluding chapter summarizes the key findings of this work and discusses their implications. An important result of this study is the suite of benchmarked model inputs that describe adequate variation in both time and space for the study domain for an entire summer season. Driven by these inputs, the model has stable performance for the entire modeling period. The model is able to reasonably reproduce observed ozone and its precursor species concentrations. HDDM-CMAQ is demonstrated to be an effective tool for determining ozone responses to precursor reductions in time and space and to delineate the relative importance of emission contributions to ozone concentrations from local versus upwind areas. Cluster analysis results demonstrate that a few ozone episodes cannot fully represent other high ozone days when both ozone production and transport are different. We conclude the chapter by providing suggestions for future research; for example, we recommend that the sensitivity study method discussed in Chapter 3 and the emission trading scenarios investigated in a companion report[1] be applied to other types of ozone episodes found in Chapter 5, to achieve a more comprehensive and effective basis for control strategy design.

[1] The parallel report conducted as part of the project focuses on the evaluation of short term and inter-basin credit trading. The report, *Impact of Short–Term, Interbasin, and Interpollutant Credit Trading on Air Quality and Credit Prices* can be found on the California Energy Commission's website.

CHAPTER 2:
Performance Evaluation of MM5-CMAQ for the SARMAP Domain

2.1 Introduction

Performance evaluation is a crucial step in air quality model applications, as it determines a model's degree of acceptability and usefulness for a specified task and establishes confidence that a model is obtaining good results for the right reasons.

Past evaluations of the Community Multiscale Air Quality (CMAQ) modeling system (Byun and Schere 2006) have focused on the eastern United States, and applications of it are usually conducted for high-ozone episodes that last for a few days. In this evaluation, the MM5-CMAQ modeling system is applied to a 15-day period from the summer 2000 Central California Ozone Study that has significant spatial and temporal variations in air quality. The middle 5-day period is a high-ozone episode, while the days before and after it are relatively clean.

Initial model diagnostic simulations are conducted for the 5-day ozone episode using the default settings determined in past studies, where focus is placed on refining model inputs and parameters according to model diagnostics, observations, and the literature. The modeling period is extended to 15 days without further changes in input parameters. Assessing model performance dynamically for these low-ozone days in the 15-day period gives more credence that the model simulates the high-ozone episodes correctly, and for the right reasons. Since ozone is a secondary pollutant, accurate predictions of ozone and NOx versus VOC sensitivity regimes depend on the success of the model's ability to reproduce precursor concentrations and their relationships with ozone.

More specifically, the model is evaluated with emphasis on following aspects: (1) spatial and temporal trends in model performance, (2) the model's ability to reproduce observed ozone and its precursor concentrations, (3) whether meteorological data assimilation significantly improves air quality predictions, and (4) the model's ability to reproduce the observed ozone sensitivity regimes. To achieve these evaluation goals, this study applies a variety of evaluation and diagnostic methods. Sensitivity analyses are conducted to diagnose discrepancies between observations and model predictions in different geographic regions, and to suggest further model improvements.

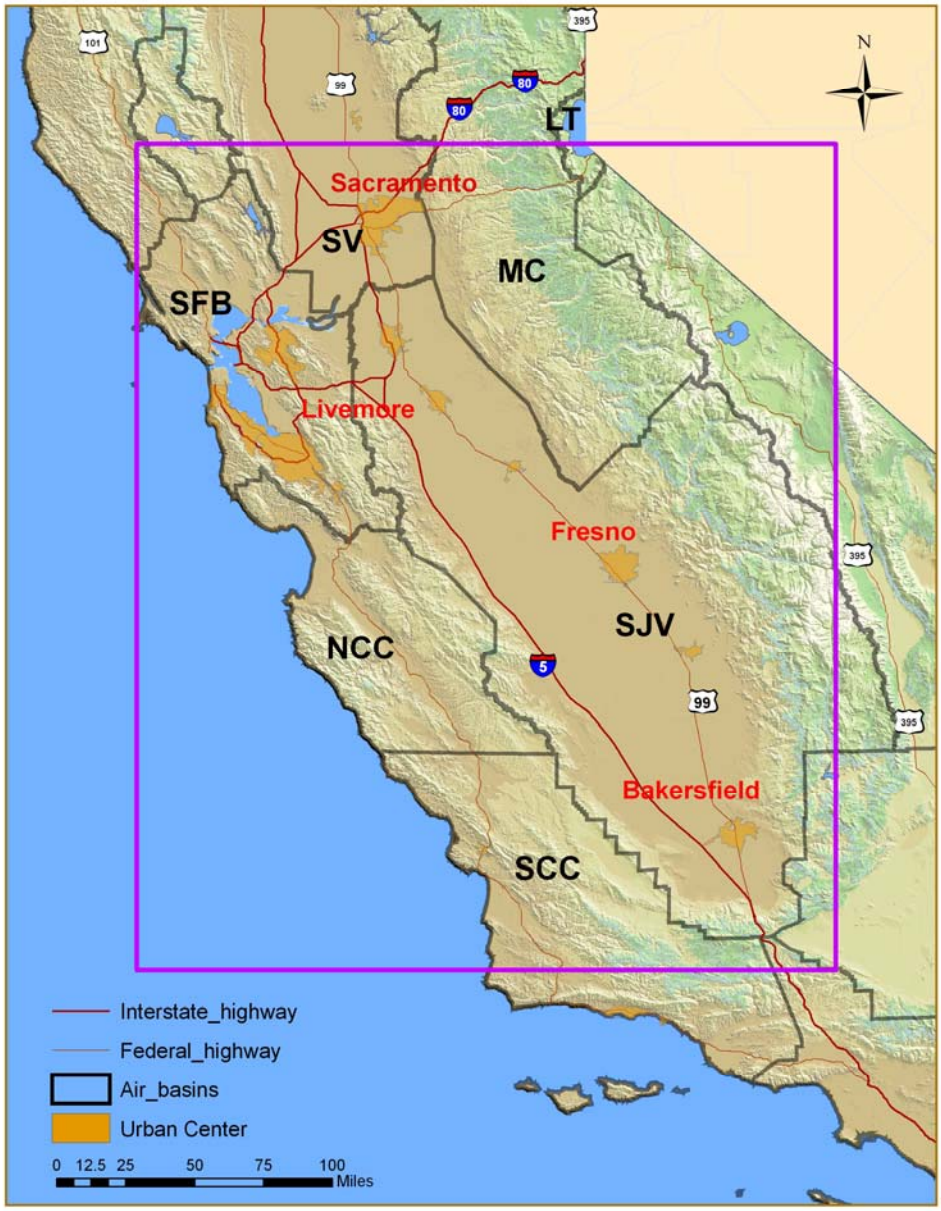
2.2 Model Description

2.2.1 Modeling Domain

The SJVAQS/AUSPEX Regional Modeling Adaptation Project (SARMAP) domain is shown in Figure 2-1. The Central Valley is surrounded by the Sierra Nevada and coastal mountain ranges. On typical summer days, westerly winds are funneled into the Central Valley through gaps in the coastal range, with large portions of the flow directed into the SJV. During the day, the flow is directed up the Sierra Nevada Mountain slopes, while at night it reverses and flows down the

slope. The San Francisco Bay area (hereinafter Bay area or SFB) and Sacramento Valley (SV) are the major upwind emission sources affecting air quality in the San Joaquin Valley. The study domain is gridded into 96 by 117 cells, with a horizontal resolution of 4 kilometers (km). Vertically, the domain is divided into 27 layers from the surface to 100 millibars (about 17 km); the near surface layers are about 20 meters (m) thick.

Figure 2-1: Modeling Domain Indicated by the Purple Rectangle



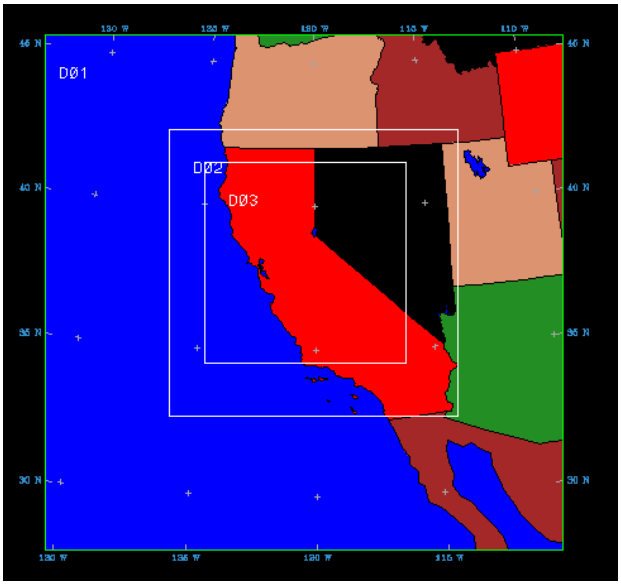
Air basins are labeled on the map for San Francisco Bay area (SFB), Sacramento Valley (SV), Mountain Counties (MC), San Joaquin Valley (SJV), and North and South Central Coast (NCC and SCC).

2.2.2 Meteorological Inputs

Gridded hourly meteorological fields are developed by J. Wilczak and coworkers at the National Oceanic and Atmospheric Administration (NOAA) with the MM5 (version 3.7) meteorological model (Grell et al. 1994), for the 15-day period of July 24 to August 8, 2000. Simulations were made with a five-layer soil Land Surface Model (NOAH LSM) and Eta Atmospheric Boundary Layer (Eta ABL) scheme that describes the surface and turbulence parameterizations. Four dimensional data assimilation (nudging) was conducted with respect to a network of surface wind observations and radar wind profilers for readings at higher altitudes operated during the Central California Ozone Study (CCOS).

Three nested grids were used: an outer 36-km resolution grid, within which was a 12-km grid, and then a 4-km grid, as shown in Figure 2-2. Vertically, the MM5 model includes 50 sigma layers with 27 layers below 1.5 km. Simulated meteorological fields with 190 by 190 grid cells with 4 km resolution were obtained for this study. The Meteorology to Chemistry Interface Processor (MCIP version 3.3) was used to construct CMAQ model-ready input files from the MM5 output. MCIP allows for consolidation of MM5 layers; 27 layers for CMAQ from the original 50 MM5 layers were used without changing the first 200 m (9 layers) in order to preserve high resolution near ground. A subset of the meteorological inputs corresponding to the SARMAP domain was used to drive air quality simulations in CMAQ.

Figure 2-2: Description of the Outer (36-km), Middle (12-km) and Inner (4-km) Nested Grids Used in the MM5 Simulation



2.2.3 Emission Inputs

Emission inputs fall into five major source categories: area, biogenic, motor vehicle, point, and forest fires. Most power generation sources, such as power plants, fall into the point sources

category. Only the smallest of distributed generation sources would be categorized as an area source. For this study, off-road mobile sources were inventoried along with area sources, rather than on-road motor vehicle emissions. The general procedure for processing these files into model-ready form is to translate and reformat the acquired files, and produce gridded output files with hourly emissions of each species in moles/second. The processing for point sources includes plume rise calculations done on a day-by-day basis to account for variations in meteorological variables.

Area Source Emissions: Area source emission estimates were obtained from the California Air Resources Board. Emissions that had been assigned to the CCOS domain (190 × 190) with hourly time resolution were speciated appropriately for SAPRC99 (Carter 2000). Estimates were day-of-week specific, with differences between weekdays (Monday–Friday) and weekends (Saturday, Sunday).

Biogenic Emissions: Date-specific biogenic VOC emissions of isoprene, monoterpenes, and methylbutenol were estimated with the BEIGIS model (Scott and Benjamin 2003). The majority of biogenic VOC emissions occur in the forested regions of the model domain, particularly in the foothills of the Sierra Nevada and coastal mountains. Isoprene emissions are the dominant daytime biogenic VOCs, with large fluxes arising from oak trees in the Sierra Nevada foothills. Additionally, certain crops are emitters of monoterpenes, leading to sources in the Central Valley. Isoprene emissions occur only in the presence of light, and increase with solar radiation until a saturation point is reached. The emissions increase exponentially with temperature up to approximately 35°C–40°C (95°F–104°F), after which they decrease.

Motor Vehicle Emissions: The motor vehicle emission inventory estimates for summer 2000 used in this research describe light-duty gasoline and heavy-duty diesel vehicle activity patterns and emissions separately by time of day and day of week. Separate activity patterns are specified for typical weekdays (Monday–Thursday), Friday, Saturday, and Sunday. Within a diurnal cycle on weekdays, diesel truck activity peaks during the middle of the day; whereas, passenger vehicles show a bimodal distribution with peaks associated with morning and afternoon commute periods. Compared to weekdays, there are large (60 to 80 percent) reductions in diesel truck traffic and the accompanying NOx emissions on weekends. Overall light-duty vehicle traffic decreases only slightly on weekends, but is shifted in time with a single broad peak in the afternoon hours. These temporal patterns in motor vehicle emissions, as well as the increasingly important role of diesel engines as a source of NOx, are described in further detail elsewhere (Harley et al. 2005). Briefly, fuel sales at the county level were combined with emission factors expressed per unit of fuel burned to estimate CO and NOx emissions. Non-Methane Organic Carbon (NMOC) emissions were estimated from CO using NMOC/CO ratios derived from ambient air measurements.

Point Source Emissions: Emission estimates were obtained from ARB. Each source is described by location, stack parameters (diameter, height, exit temperature, exit velocity), and hourly emissions of pollutants speciated into SAPRC99. Emissions are day-of-week specific, with differences between weekdays (Monday–Friday), Saturdays, and Sundays. Plume rise extracted

from the emission preprocessor SMOKE was used to estimate emissions injected by each point source into various layers, using time and location-specific meteorological variables from MM5.

Forest Fire Emissions: These emission estimates were obtained from ARB for the five-day ozone episode, July 29 to August 2, 2000, in a vertically distributed form. Each fire is represented by a number of individual point sources, each point source representing the emissions at a particular height, after a buoyancy calculation has been applied. The processing procedure for estimating the fire parameters and the plume rise is described in FEJF (1996) and Loomis (2003). The NOx and CO emission profiles associated with area, motor vehicle, point, and fires for days 211 through 216 are illustrated in Figures 2-3, 2-4, and 2-5. Since fire emissions are only available for a 5-day period, for the purpose of consistency they are not included in any 15-day simulations. In further work described later in this report, fire emissions for the entire season are included.

Emissions files from the various source categories were processed and merged to produce an input file for CMAQ simulations from July 24 to August 8, 2000, and are also used for five-day diagnostic simulations for Intensive Observational Period 2 (IOP2). Sample figures showing spatial and temporal patterns of emissions are presented below. The fire inventory is for days 211–216.

Figure 2-3: NOx Emission Profiles Summed over the Domain for the Various Categories

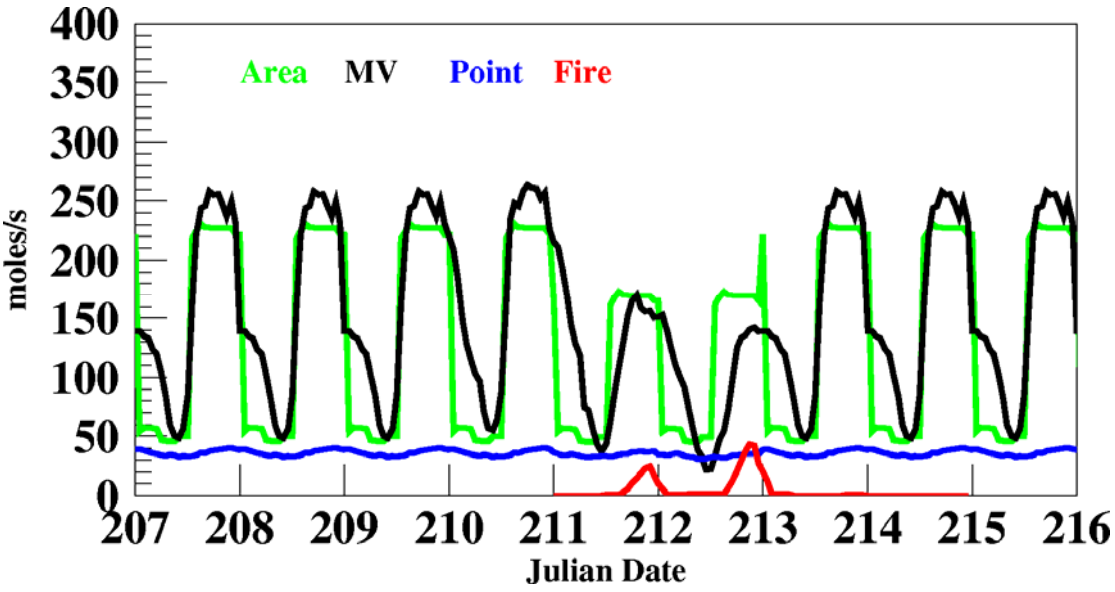


Figure 2-4: CO Emission Profiles Summed over the Domain for the Various Categories

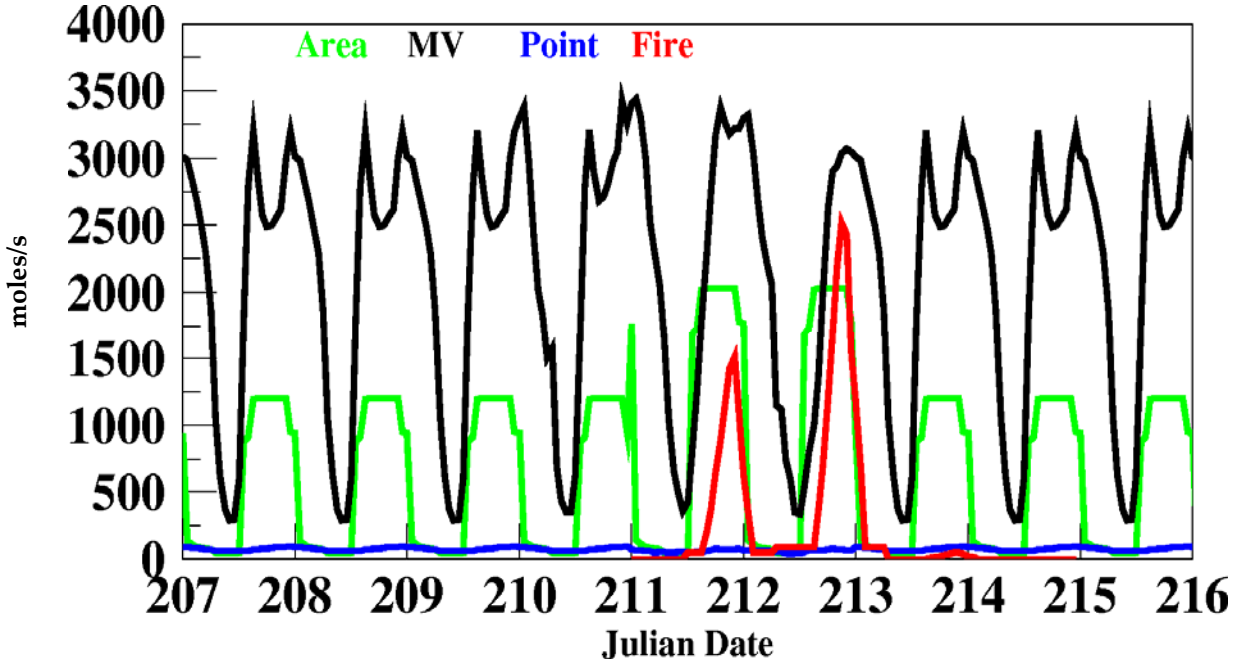
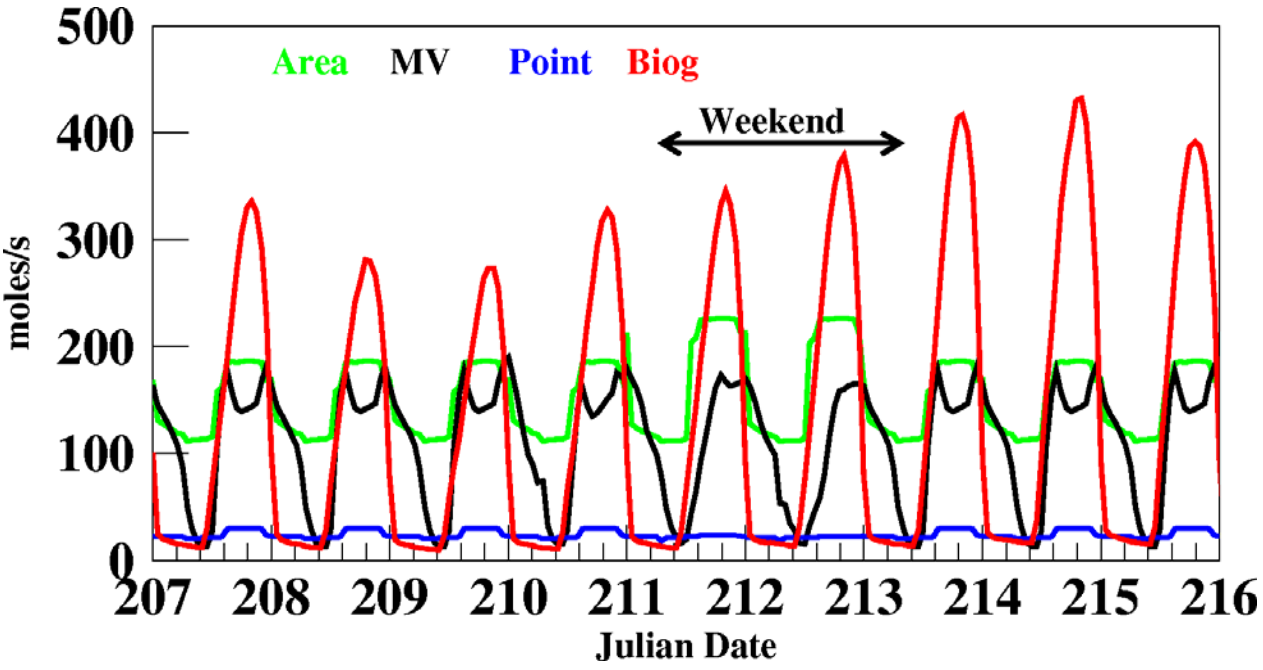
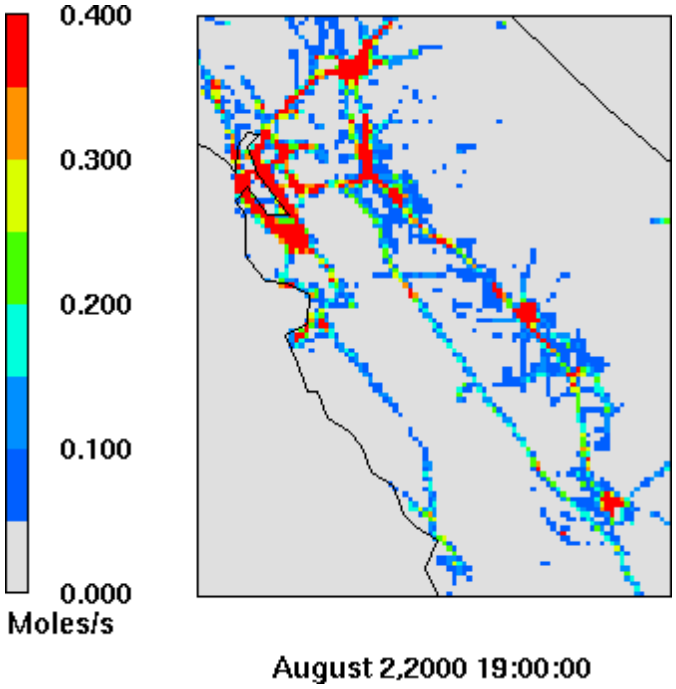


Figure 2-5: VOC Emission Profiles Summed over the Domain for the Various Categories



Processed emissions of NOx for the study domain are shown in Figure 2-6.

Figure 2-6: Spatial Distribution of NOx Emissions at Noon PDT



The urban centers of the SF Bay area, Sacramento, and Fresno can be seen. Also prominent, extending in a NW to SE direction, are highways 101, I-5 and 99 going from left to right, respectively.

2.2.4 Community Multi-scale Air Quality Model

The CMAQ model (Byun and Schere 2006) incorporates up-to-date representations of atmospheric chemistry and physics, is peer-reviewed, and is publicly available. It is a multi-scale, multi-pollutant model that has been evaluated for numerous ozone and particulate matter (PM) studies, most of which are focused on the eastern parts of the United States. The atmospheric transport and chemistry in CMAQ are described by the Atmospheric Diffusion Equation (ADE):

$$\frac{\partial C_i}{\partial t} = -\nabla \cdot (\mathbf{u} C_i) + \nabla \cdot (\mathbf{K} \nabla C_i) + R_i + E_i$$
$$(i = 1, 2, \cdots, N) \qquad (2\text{-}1)$$

where u is the three-dimensional wind field, K is the turbulent diffusivity tensor, and N is the total number of chemical species. *Ci, Ei,* and *Ri* are the grid cell average concentration, emission rate, and net chemical formation of species *i,* respectively. Various chemical mechanisms are included in CMAQ: the Carbon Bond (CB-IV and CB-V), Regional Acid Deposition (RADM), and SAPRC97/99 (Carter 2000). For gas-phase problems, the diffusion equations are numerically integrated with the Sparse Matrix Vectorized Gear (SMVGEAR) and the Euler Backward Iterative (EBI) ordinary differential equation solvers. Advective transport is treated with the

piecewise parabolic method. Useful features such as the Decoupled Direct Method sensitivity analysis module and Process Analysis module contained in CMAQ can be used to quantify pollutant sensitivities to input parameters and the contributions of individual physical processes and chemical reactions to the predicted concentrations of chemical species.

2.2.5 Model Evaluation and Diagnostic Methods

Model evaluation in this study is based on hourly data from observational sites and model grid points in the SARMAP domain. Different evaluation and diagnostic tools address model performance from various perspectives. The evaluation and diagnostic methods employed in this study are summarized in Table 2-1.

Table 2-1 Evaluation and Diagnostic Tools

Method	Usage
Pattern Statistics	Describe how well patterns in modeled and observed values match each other.
Mean Bias	Average sign in model prediction errors
RMSE, Gross Error	Magnitude of model prediction errors
Spatial Kriging	Spatial variation in model performance
Ozone Production Efficiency	Compare limiting regimes presented in the observed and modeled air masses
Sensitivity Analysis	Identify the influential parameters that affect model performances and verify reasonable model responses

Pattern statistics compare variability captured in the modeled data with that in observations (e.g., correlations between modeled and observed data). This type of statistics do not indicate overprediction versus underprediction, but rather compares the pattern similarity between observation and model predictions. Taylor has devised a very useful diagrammatic form (termed "Taylor diagram") to visualize three pattern statistics on the same two-dimensional plot (Taylor 2001). Considering a reference field (r) and a modeled field (f), the three pattern statistics are:

Standard deviation of the reference field or the modeled field (σ_r or σ_f)

Correlation between the two fields R

$$R = \frac{\frac{1}{N}\sum_{n=1}^{N}(f_n - \bar{f})(r_n - \bar{r})}{\sigma_f \sigma_r}$$

Root mean square difference E'

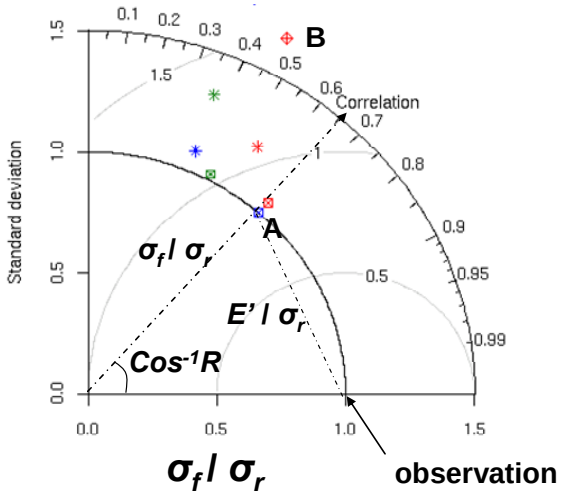
$$E' = \left\{ \frac{1}{N} \sum_{n=1}^{N} \left[(f_n - \bar{f}) - (r_n - \bar{r}) \right]^2 \right\}^{1/2}$$

These three pattern statistics satisfy the Law of Cosines:

$$E'^2 = \sigma_f{}^2 + \sigma_r{}^2 - 2\sigma_f \sigma_r R$$

and therefore can be placed on the same two-dimensional plot (e.g., see Figure 2-7).

Figure 2-7: Taylor Diagram Example



The diagram shows lines of constant correlation, grey arcs of constant E', and black arcs of constant standard deviation. E' and standard deviation are normalized by standard deviation of observations so that observations are always at (1,0).

Modeled field A has the same standard deviation as the observed data, while model field B has a larger standard deviation, and therefore varies more than observed. Similarly, it can be seen that A correlates better with observations and has a smaller root-mean-square difference, and thus better resembles the observed pattern.

Mean bias and gross error indicate the sign and magnitude of model prediction errors, respectively. These statistics are recommend (U.S. EPA 1991) for evaluating urban-scale ozone models.

While comparisons between model and observations are made at individual measurement sites, a more comprehensive spatial context can be achieved by interpolating these point biases (or errors) into a continuous surface through kriging, as practiced in Mena-Carrasco et al. (2007). *Kriging* is a geospatial interpolation technique to estimate point bias at an unobserved location by linearly combining values at observed sites (Cressie 1993). The point value is estimated by an additive model, consisting of an average term, a spatially correlated term, and a random error term (Oliver and Webster 1990). Assuming spatial isotropy (meaning covariance between any two points is a function of only the distance between them), the spatially correlated component is characterized by a covariance function; in this case, Matern family (Matern 1960), where covariance between any two points is related only to the distance between them. Kriging has been widely used in interpolating surface measurements to generate maps of air pollution for health studies to estimate exposure (Wackernagel et al. 2004; Liao et al. 2006). In addition to generating a continuous surface, kriging has been used to address incommensurateness issues in model-observation comparison in Swall and Davis (2006) for spatially slow varying sulfate concentrations.

Besides examining individual species predictions, a previous study related odd oxygen Ox to NOz (all compounds that are products of the atmospheric oxidation of NOx) in terms of ozone production efficiency (OPE) (Trainer et al. 1993). *Ozone production efficiency* is defined here as the slope of the curve Ox versus NOz (Mena-Carrasco et al. 2007), which measures how much odd oxygen is formed per unit conversion of reactive nitrogen species NOx to NOz. Ozone production efficiency reflects ozone sensitivity to NOx in the air mass of interest, and thus can be used to evaluate how well modeled ozone sensitivities match the observed ones.

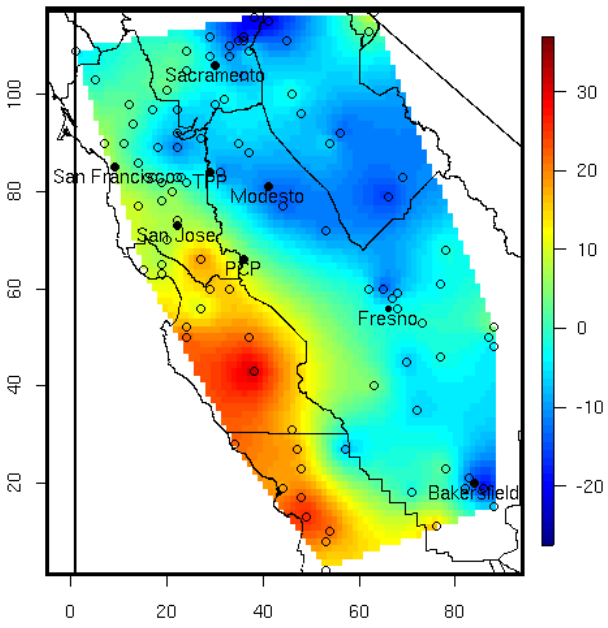
Sensitivity analysis can be used to determine how concentrations of certain species change when an input parameter is perturbed, and thus can be used to identify influential input parameters. This can be done with brute force perturbations or with the Decoupled Direct Method (DDM) (Dunker 1981; Yang et al. 1997; Dunker et al. 2002).

2.3 Initial Simulations and Model Improvements

2.3.1 Initial Simulations and Diagnostics

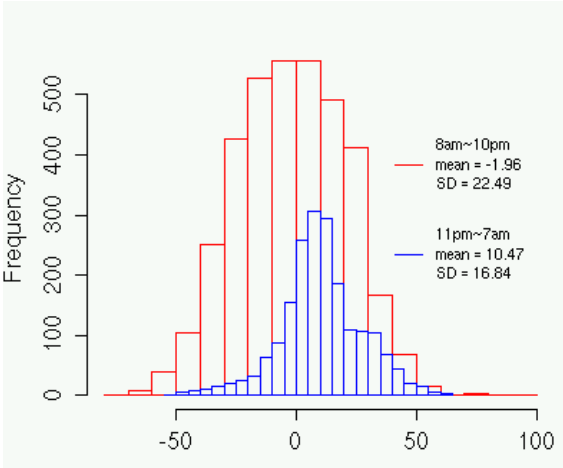
Initial simulations are conducted for the five-day ozone episode (July 29 to August 2) with CMAQ V4.5 configured with the SARPRC 99 chemical mechanism, SMVGEAR (sparse matrix vector gear code) solver, and the piecewise parabolic advection scheme. Boundary conditions are set to default values, with ozone boundary conditions constant at 40 ppb for all layers and all four boundaries. Positive biases are found among the coastal sites, while slightly negative biases are seen inland (Figure 2-8). The model is also found to overpredict ozone concentrations at night (Figure 2-9).

Figure 2-8: Kriged Daytime Ozone Normalized Biases (%)



(cutoff = 40 ppb), with unnudged meteorological inputs. Measurement sites are marked by hollow points. Positive biases are seen in the North and South Central Coasts, while the Mountain Counties and Sacramento Valley exhibit negative biases.

Figure 2-9: Ozone Biases (ppb) Histograms (Day and Night)



O_3 biases (ppb)

A series of diagnostic simulations (Table 2-2) are conducted to investigate effects of input parameters focusing on boundary conditions (1 and 2), vertical diffusion (3), ozone dry deposition rate (4), and ozone titration effects at night (5). The mostly affected species and areas are summarized in the last column of Table 2-2. One can see that all of these parameters considered in simulation numbers 1 to 4 mostly affect pollutant predictions in the coastal areas and/or nighttime ozone concentrations. Ozone boundary conditions are especially important for correctly simulating ozone concentrations along the coast.

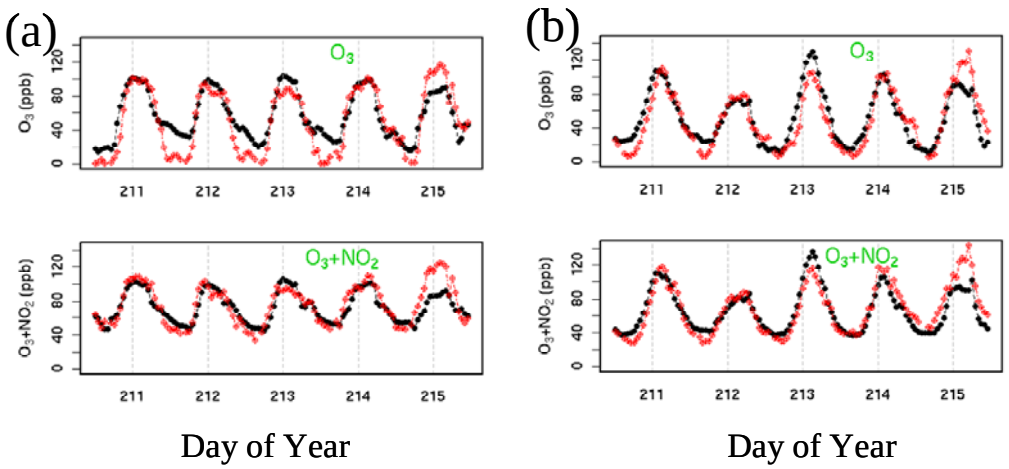
The model was found to overestimate ozone concentrations at night at many locations. The positive biases in ozone predictions are often seen at stations that are close to roadways or other large sources of NOx. As an example, Figure 2-10 shows two stations with different emission settings: the Bakersfield station is close to mobile NOx sources (Figure 2-10a), while the Turlock station is not (Figure 2-10b). The model overpredicts nighttime ozone at the Bakersfield station but not at Turlock. The observed ozone is a point measurement, and can be decreased substantially by the titrating reaction with nearby NO emissions to form NO_2 ($O_3 + NO \rightarrow NO_2 + O_2$) at night when turbulent mixing is reduced. In contrast, the model distributes emissions in a 4-km by 4-km grid cell, and therefore predicted ozone cannot capture the sub-grid variabilities seen at a single point. As the level of odd oxygen (O_3+NO_2) is not changed by ozone titration, it is a more conserved quantity and is better for evaluating model performance at night. When the titration effect is considered (i.e., comparing NO_2+O_3), the nighttime positive biases in the model are reduced (Fig 2-10a). Nighttime overprediction is reduced from 10.5 ppb to 3.5 ppb after titration effects are accounted for. The unexplained biases may be due to inaccurate allocation of emission rates at night.

Table 2-2: Summary of Initial Model Diagnostic Simulations

No.	Simulation ID	Process or Parameter Adjusted	Most Affected Species / Geographic Areas
1	O_3 Sensitivity to Boundary Conditions (BC)	Ozone sensitivity calculated by the DDM method to boundary ozone, and its precursors (NO, NO_2, and VOC). Brute force perturbation at different boundary layers.	Simulated ozone is most sensitive to boundary ozone. Coastal sites are most affected. Perturbation at various layers affects nearby local layers the most.
2	Constant BC vs. Vertically Varying BC	Change constant ozone boundary (40 ppb at all layers) to a vertically varying profile (20 ppb at bottom, and increasing aloft) derived from Trinidad head measurements.	A decrease in modeled coastal ozone, and an increase in inland ozone, by ~5 ppb.

No.	Simulation ID	Process or Parameter Adjusted	Most Affected Species / Geographic Areas
3	Minimum Eddy Diffusivity (Kzmin)	Change the default value of minimal vertical eddy diffusivity used in CMAQ from 0.5 to 0.1 square meters per second (m^2s^{-1}) based on Jacobson (1999).	Lowered nighttime and some daytime morning O_3 of the inland areas by ~10 ppb. Both day- and night-time coastal ozone are affected.
4	Ozone dry deposition rate (Vd_O3)	Brute force sensitivity conducted by adding 0.05 centimeters per second (cm/s), 0.1 cm/s, and 0.5 cm/s to the default ozone dry deposition rate in the MM5 input file. Change the zero ozone dry deposition rate to 0.04 cm/s based on Wesley (2000) and measurements by Faloona (2006).	Nighttime ozone and ozone at coastal sites are most sensitive to ozone dry deposition rate.
5	Nighttime Titration	Compare simulated NO_2+O_3 to observed values.	Sites near roadways, where $NO+O_3 \rightarrow NO_2 + O_2$ occurs. Improved agreement during nighttime.

Figure 2-10: Ozone and Odd Oxygen (O_3 + NO_2) Time Series Plots



Shown for (a) the Bakersfield station, close to roadways, and (b) the Turlock station, in a residential area. The dashed vertical lines indicate mid-day. Observations are in red and modeled values in black.

2.3.3 Refining Input Parameters

Input parameters are refined according to the diagnostic simulation results, field measurements, and literature, and this results in the "base case" inputs. Ozone dry deposition velocity over the ocean is increased from the CMAQ default value of zero to 0.04 cm/s according to measured values (Faloona 2006, personal communication). Minimum eddy diffusivity in CMAQ (Kzmin = 0.5 m^2/s) appears too high for stable marine layers on the western boundary, and this causes excessive vertical mixing and increases surface ozone. Assuming wind speeds of about 10 m/s and surface roughness lengths for the ocean to be between the suggested values of 1.5×10^{-5} and 1.5×10^{-3} m, we calculated Kz to be between 0.0015 and 0.15 m^2/s. Kzmin is thus set to 0.1 m^2/s.

Constant pollutant concentrations are set for each of the four lateral boundaries of the domain, with the western boundary treated differently from the others (Table 2-3). The western inflow boundary is mostly over the ocean, and therefore its chemical species are set to clean marine background concentrations. Vertically varying O_3 is based on averaged August ozone sonde measurements made at Trinidad Head, California (Newchurch et al. 2003). Nitrogenous species (NO: 0.01 ppb; NO_2: 0.03 ppb; etc.) and a suite of VOC species take values measured in the marine background free troposphere (Nowak 2004). The other three boundaries are dominated by outflows, and the same boundary values used in past studies conducted of this domain by ARB (Kaduwela 2006) are applied here.

Table 2-3: Boundary Conditions for Selected SAPRC-99 Species

SAPRC-99 Species	Clean Western Boundary (ppb)	North/East/South Boundary (ppb) *
CO	120 (Goldstein et al. 2004)	200
NO	0.01 (Nowak et al. 2004)	0.05
NO_2	0.03 (Nowak et al. 2004)	1
PAN	0.15 (Nowak et al. 2004)	0.005
PAN2	0.02 (Nowak et al. 2004)	0
O_3 at sfc	22 (Newchurch et al. 2003)	40
O_3 at 1 km	44 (Newchurch et al. 2003)	48
O_3 at 2 km	60 (Newchurch et al. 2003)	60
O_3 at 3 km	60 (Newchurch et al. 2003)	60
HCHO	0.3 (Singh et al. 2001)	2
CCHO	0.15 (Singh et al. 2001)	0.456

SAPRC-99 Species	Clean Western Boundary (ppb)	North/East/South Boundary (ppb) *
ACETONE	0.89 (Nowak et al. 2004)	1
ALKANE-1	1.06 (Nowak et al. 2004)	10
ALKANE-2	0.13 (Nowak et al. 2004)	2.5
AROMATIC-1	0.019 (Nowak et al. 2004)	0.35
AROMATIC-2	0	0.25
ISOPRENE	0	0.1
OLEFIN-1	0	0.5
OLEFIN-2	0	0.2

Notes: PAN = peroxyacetyl nitrate; HCHO = formaldehyde; CCHO = other aldehydes.

* Based on personal communication with Ajith Kaduwela (ARB 2006).

Table 2-4 presents a comparison of model performance in simulating 1-h and 8-h peak ozone under old and improved input parameters. All the evaluation statistics indicate better agreement between modeled and observed peak ozone. Improvements are seen in the SJV and SV according to all the statistics, but not the SFB, which will be investigated in later sections with sensitivity analysis.

Table 2-4: Peak Ozone (1-h and 8-h) Predictions Under Old and Improved (Base Case) Settings

8-h Peak Ozone	All Air Basins		SJV		SV		SFB	
	Old Settings	Base Case	Old Settings	Base Case	Old Settings	Base Case	Old Settings	Base Case
Mean Bias (ppb)	-14	-5	-18	-8	-23	-13	0	4
RMSE (ppb)	22	18	22	15	27	24	17	16
N. Bias (%)	-14	-4	-20	-8	-26	-13	5	11
N. Gross Err (%)	22	18	21	14	28	23	22	20

8-h Peak Ozone	All Air Basins		SJV		SV		SFB	
	Old Settings	**Base Case**	**Old Settings**	**Base Case**	**Old Settings**	**Base Case**	**Old Settings**	**Base Case**
Sample Size	332	332	122	122	72	72	62	62
1-h Peak Ozone	All Air Basins		SJV		SV		SFB	
	Old Settings	Base Case	Old Settings	Base Case	Old Settings	Base Case	Old Settings	Base Case
Mean Bias (ppb)	-15	-5	-21	-9	-26	-13	5	13
RMSE (ppb)	28	25	28	21	33	29	24	29
N. Bias (%)	-13	-2	-19	-7	-26	-11	11	21
N. Gross Err (%)	25	22	23	16	28	23	26	33
Sample Size	295	295	109	109	62	62	71	71

Cutoff = 40 ppb. Performances at all the sites, as well as individual air basins of the San Joaquin Valley (SJV), Sacramento Valley (SV), and San Francisco Bay area (SFB) are shown.

2.4 Evaluation of Base Case Simulations

Base case simulations are conducted with input parameters determined from Section 2.3 for the 15-day period. Additional simulations for the middle 5-day ozone episode are performed using nudged meteorological fields to investigate the effects of wind nudging.

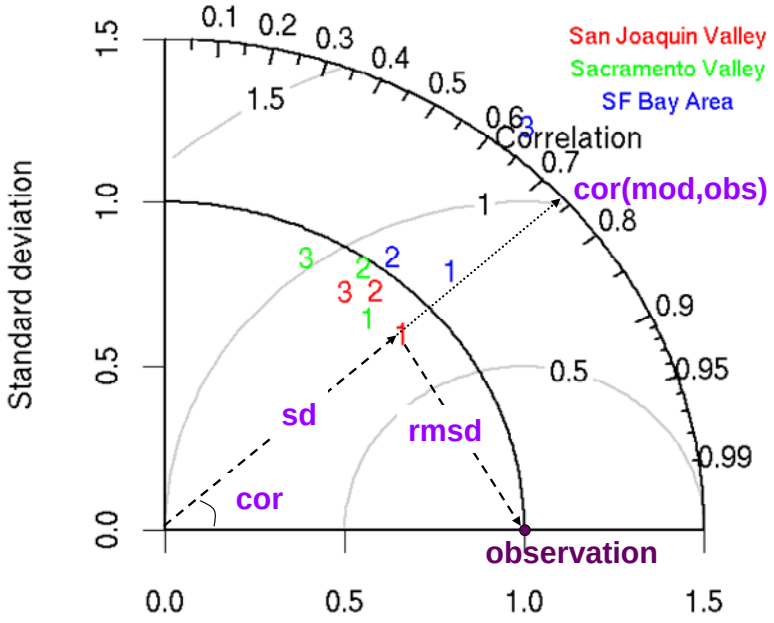
2.4.1 Spatial and Temporal Trends in Model Performance

The model's ability to reproduce the variation and correct levels of observed ozone is examined with a Taylor diagram (Taylor 2001) and normalized biases. Daytime comparison statistics are reported with respect to ozone. At night, when strong convective mixing is absent, observed ozone at measurement sites is often more affected by titration with NO emitted at nearby roadways. Hence we compare odd oxygen Ox (NO_2+O_3) concentrations to evaluate model performance at night. Carbon monoxide and non-methane hydrocarbons (NMHC) are also compared between model and observations without distinguishing day and night.

On a Taylor diagram the correlation coefficient and the root-mean-square difference (RMSD) between the modeled and observed values, along with the ratio of the standard deviations of the two patterns, are all indicated by a single point on a two-dimensional plot (Figure 2-11).

Across the three air basins that have the highest ozone levels, model simulation of the middle five-day ozone episode shows a consistently better match to the observed standard deviation, while the first five-day period shows consistently better linear correlations and root-mean-square differences across the three air basins. In general, points representing modeled patterns in different air basins and time periods are clustered together, indicating similar model performance across time and space. All the pattern metrics indicate that model performance in the San Francisco Bay area during the last five-day period needs improvements relative to the first two five-day periods.

Figure 2-11: Taylor Diagram Showing Daytime Pattern Statistics

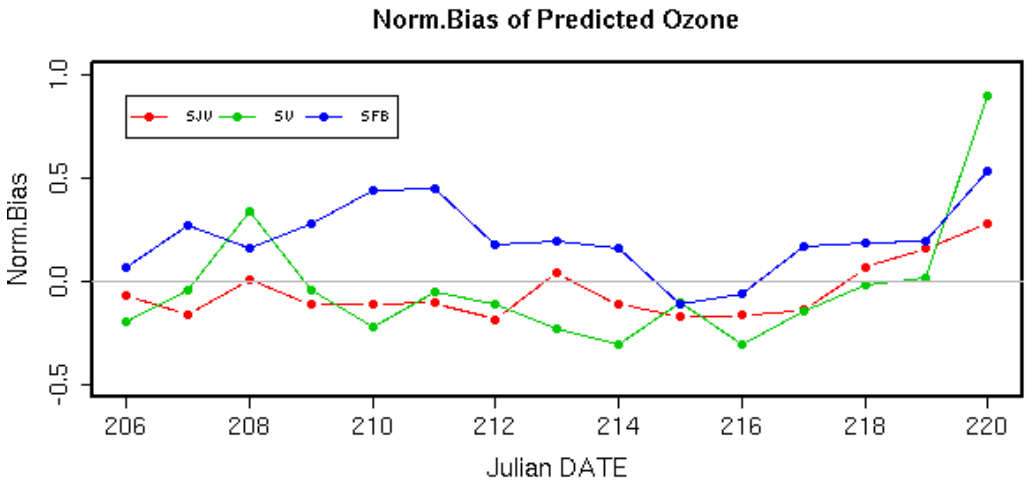


Normalized by observed standard deviation, for July 24–29 (1), July 29–August 2 (2), August 2–7 (3), respectively, for three air basins. Note that once normalized by the observed standard deviation, the observation is always plotted at polar coordinate (1, 1). cor: correlation between observed and modeled ozone; sd: standard deviation; rmsd: root-mean-square difference. Being closer to "observation" point (sd→1, and cor→1, rmse→0) indicates better model performance.

Pattern statistics are useful for comparing the anomalies in two data sets, while normalized mean bias indicates either overprediction or underprediction. Figure 2-12 indicates that the model slightly under-predicts daytime ozone in the San Joaquin Valley and Sacramento Valley in general, but over-predicts ozone in the Bay area, especially for the first half of the simulation period. There is no clear temporal trend in model biases, but we do observe overprediction on

the last day, and this corresponds to the positive bias present in MM5 temperature fields during that day, reported by Bao et al. (2006).

Figure 2-12: Daytime Normalized Biases at Three Air Basins



Cutoff value = 40 ppb.

Table 2-5 shows evaluation metrics for modeled peak ozone, daytime NOx, nighttime Ox, as well as CO and NMHC across the three five-day periods. Peak ozone predictions are consistent through the three five-day periods, especially for 8-h peak ozone. The model produce acceptable performance in simulating peak ozone concentrations in the high-ozone episode as well as for the relatively clean days. This is in accord with U.S. EPA guidelines that suggest acceptable model performance with normalized biases within $\pm$ 15 percent, and normalized gross error less than 30 percent. Overall, CMAQ under-predicts peak ozone concentrations, which may arise from comparing grid averaged values with point referenced measurements.

Gross errors suggest CMAQ does a better job in predicting peak ozone concentrations than the precursor concentrations. In addition to model deficiencies, this may be attributed to errors in measured precursor concentrations in a less densely covered measurement network relative to ozone. Chemiluminescent analyzers used to measure NOx also respond to other nitrogen-containing air pollutants; CO measurements are rounded to parts per million (ppm) instead of ppb, with a loss of precision; NMHC measurements are sparse, both spatially and temporally, which bears on representativeness issues. In addition, CMAQ volatile organic compound species (VOCs) are highly lumped according to reactivities, instead of carbon numbers or whether they are oxygenated or not. Therefore, the mapping of NMHC species from CMAQ outputs to observations is subject to several potential errors. As a result, one does not expect precursor-prediction performance to be as good as ozone-prediction performance.

Table 2-5: Summary Statistics for Three 5-Day Periods

	Normalized Bias (%)			Normalized Gross Error (%)		
	1	2	3	1	2	3
1-h Peak O_3	-10	-7	-4	17	20	18
8-h Peak O_3	-15	-12	-11	17	17	18
Daytime NOx	0	-4	3	57	60	64
Nighttime Ox	8	10	8	19	29	22
CO	-33	-34	-50	43	45	56
NMHC	6	-3	NA	64	49	NA

Peak ozone cutoff value: 60 ppb, NOx and Ox: 10 ppb, CO: 100 ppb, NMHC: 30 ppb Carbon.

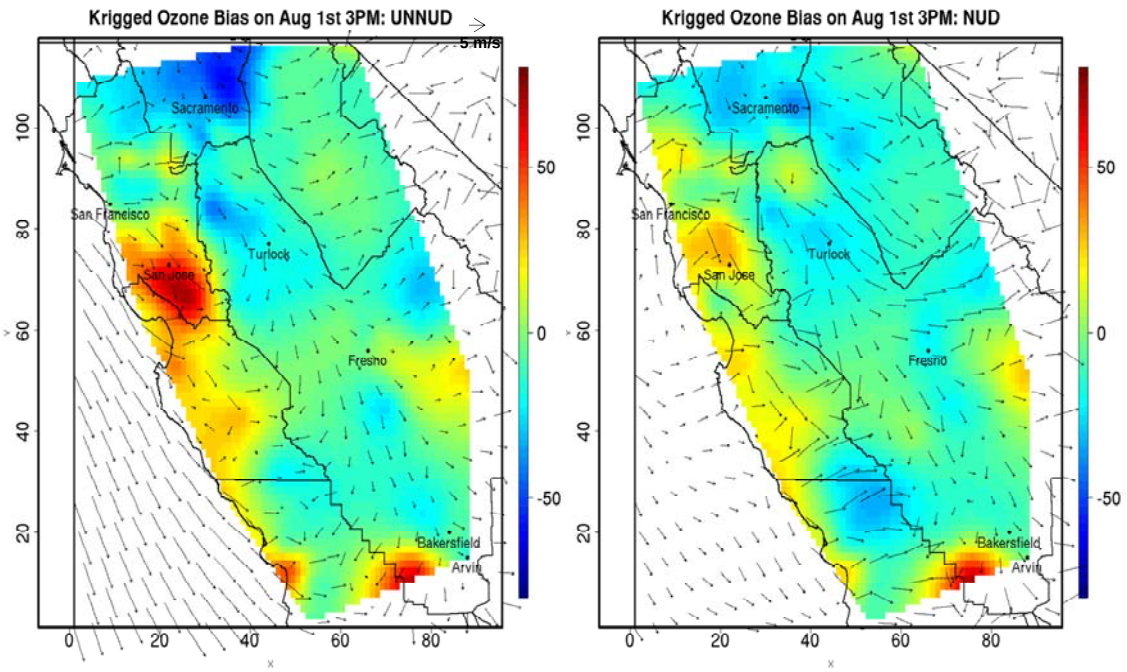
July 24–29 (1), July 29–August 2 (2), August 2–7 (3).

2.4.2 Effects of Meteorology Nudging on Air Quality Simulations

Four-dimensional data assimilation provides clear improvements in MM5 simulated wind fields in terms of wind speed and direction at all altitudes. The largest improvements appear on Day 214 (Bao et al. 2006), when we also see the largest improvements in spatial ozone biases (Figure 2-13).

Improved wind speed and direction ensure better representation of transport in the model, which leads to more correct localization of ozone plumes. Ozone biases on Day 214 are kriged to generate a continuous error surface, and are shown at 3 p.m. local time, when ozone levels are the highest. Before nudging, modeled meteorology appears to have weaker westerly flows, which prevents the Bay area ozone plume from entering the San Joaquin Valley, causing overprediction in the Bay area and underprediction in the northern San Joaquin Valley. Another significant improvement occurs in the Sacramento Valley. Before nudging, the southward winds appear to be too strong in the Sacramento Valley, resulting in a shorter residence time for local pollutants to accumulate. Thus, a large underprediction of ozone is seen in Sacramento. After observational nudging, ozone biases are reduced by about a factor of two.

Figure 2-13: Spatial Distribution of Ozone Biases (%) at 3 p.m. on Day 214



Left: with unnudged winds; right: with nudged winds. Surface winds averaged from 1–3 p.m. are plotted. Labels along the domain are 4-km grid cell indices.

Although effects of wind nudging on particular days are large, overall effects are not distinguishable in both ozone and its precursor predictions (Table 2-6).

Table 2-6: Normalized Bias and Gross Error for Model Simulations

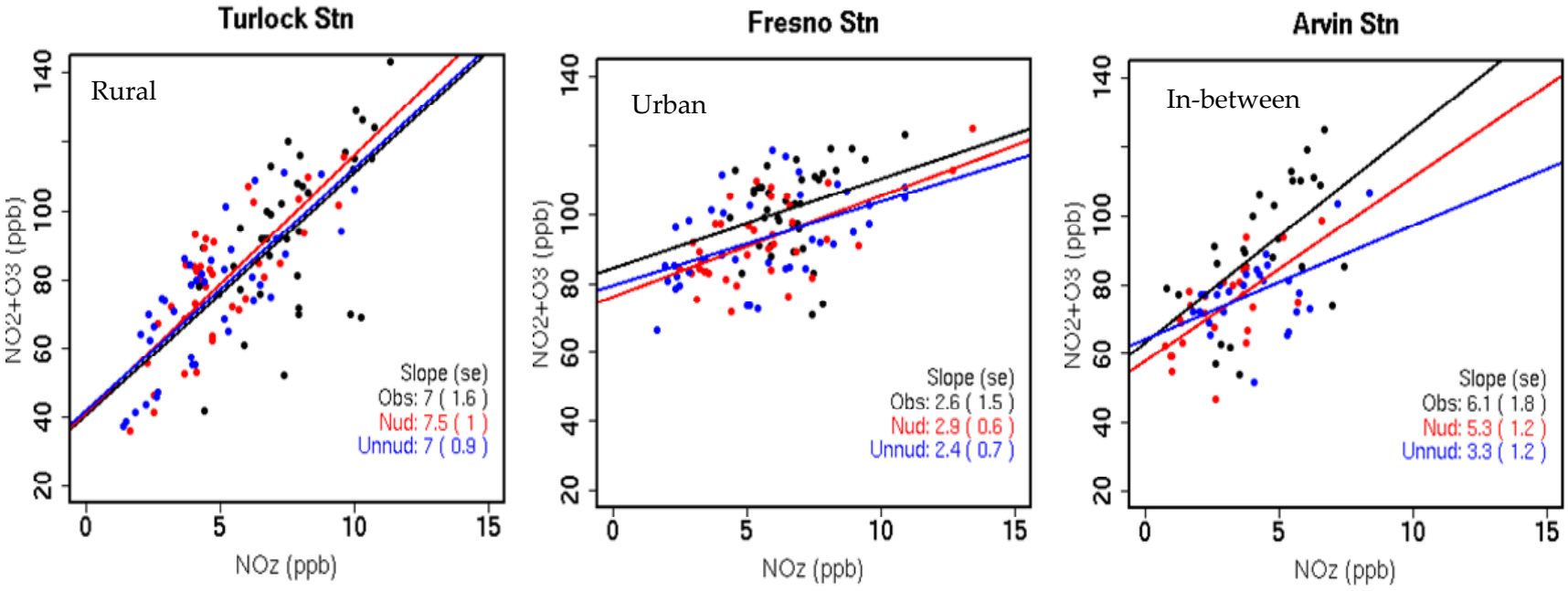
	Normalized Bias (%)		Normalized Gross Error (%)	
	Unnud	Nud	Unnud	Nud
1-h peak O_3	-7	-11	20	16
1-h peak O_3	-12	-14	17	17
Daytime NOx	-4	-4	60	53
Nighttime Ox	10	14	29	29

2.4.3 Ozone Production Efficiency

Model performance has so far been examined in terms of predicting individual species concentrations. Modeled versus observed ozone production efficiencies (OPEs, ozone produced per unit of precursor consumed), defined in Section 2.2.5, are compared here to evaluate how well modeled ozone sensitivities match the observed ones. Sources of discrepancy between

modeled and observed OPE at a particular location and time include emission uncertainties, errors in the representation of ozone chemistry, temperature errors, and misplacement of ozone plumes by transport errors. Only three sites have valid daytime NOz observations available, and weekday OPEs in the middle five-day period are presented for these sites (Figure 2-14). Fortunately, ozone chemistry at these sites represents three typical air masses in terms of their distances from the major nearby emission sources. Turlock station is located in the northern San Joaquin Valley, downwind of Bay area emissions, and often experiences aged air masses (NOx poor) transported from the Bay area. Fresno is right at the urban emission center in the middle San Joaquin Valley, representing air masses of fresh emissions rich in NOx. Arvin station in the southern San Joaquin Valley is at a moderate distance downwind of Bakersfield emissions. Ozone production efficiency is considered separately for modeled data driven by meteorology with and without four-dimensional data assimilation to investigate the effects of transport errors on predicted OPEs. In Figure 2-14, we see that modeled OPEs with or without wind nudging are both very close to the observed values at Turlock and Fresno. A large OPE at Turlock indicates efficient ozone production, typical of NOx-limiting conditions in aged air masses. Ozone production efficiency at Fresno is much lower, which indicates that the air mass is very close to the emission sources. Observed OPE (6.1) at Arvin indicates NOx-limiting conditions, while modeled OPE (3.3) without wind nudging implies that air masses are less aged. Wind nudging brings modeled OPE from 3.3 to 5.3; much closer to the observed value. Since the first two stations are either far away from or very close to the emission sources, ozone sensitivities to NOx in the air masses at each are both less affected by transport errors. In contrast, wind direction and speed determine how much Arvin is influenced by its nearby upwind emissions in Bakersfield. Similar modeled and observed OPEs across all three sites are seen when nudged wind fields are used, suggesting that the model provide a good representation of ozone sensitivities when transport errors are minimized.

Figure 2-14: Ozone Production Efficiency (the Linearly Fitted Slope) along with Standard Errors (se)



Three sites evaluated for observed and modeled (with and without wind nudging) daytime data during weekdays in the middle five-day period.

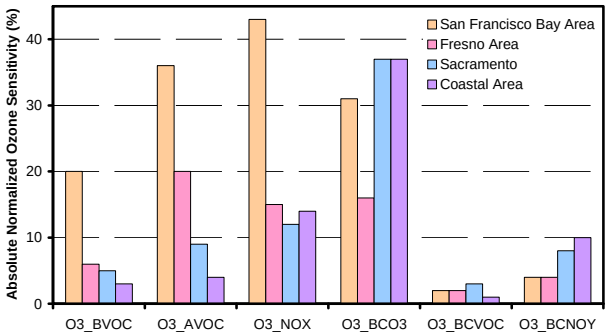
2.5 Final Sensitivity Analysis and Discussions

According to the previously discussed evaluations, model predictions in the SJV generally agree well with observations, consistently across both high ozone and low ozone days. Moreover, the overall model performance is not significantly improved by nudging the input meteorology fields toward observations, despite significant error reduction at certain times and locations. Ozone production efficiency examined for the three sites in the SJV also suggests that the model adequately captures the limiting regimes of the air masses in typical urban and rural settings. For air masses between the urban and rural settings, ozone regimes predicted by the model are sensitive to transport errors.

The CMAQ model generally overpredicts ozone along the coast. Model performance in the SFB air basin is less satisfactory relative to that in the SJV and the SV, in terms of matching the observed pattern and capturing the correct ozone and precursor levels across the 15-day period. Simulated pollutant levels in the SFB can be greatly affected by uncertainties in the wind fields as seen in Figure 2-13. The large emission sources and proximity to the western boundary of the SFB also subject it to the effects of uncertainties in the emission inventories and boundary conditions.

Ozone sensitivities to boundary conditions (ozone, VOC, nitrogen oxides), and emissions (NOx, anthropogenic VOCs [AVOC], and biogenic VOCs [BVOC]) are calculated with the Decoupled Direct Method. The mean absolute fully normalized sensitivity coefficients are plotted for different geographical areas (Figure 2-15). The coastal areas are most sensitive to NOx emissions and boundary ozone conditions. The San Francisco Bay area is sensitive to both anthropogenic and biogenic emissions, as well as boundary ozone concentrations. The Fresno area in the SJV is most sensitive to anthropogenic emissions, but less sensitive than the SFB. Sacramento is most sensitive to boundary conditions; especially ozone, due to its proximity to the northern boundary.

Figure 2-15: Mean Absolute Fully Normalized Sensitivity Coefficients of Ozone



Sensitivities to emissions of biogenic VOC (O3_BVOC), anthropogenic VOC (O3_AVOC), NOx (O3_NOx), boundary O3 (O3_BCO3), boundary VOC (O3_BCVOC), and boundary NOy (O3_NOy).

Besides the wind speed and direction of the incoming westerlies, the Bay area ozone levels are also highly sensitive to uncertainties in all the emission sources, as well as boundary conditions, which suggests great challenges for correctly simulating ozone concentrations in this area.

CHAPTER 3:
Sensitivity Analysis of Ozone Formation and Transport[2]

3.1 Introduction

Ozone is a secondary pollutant, and a key task in its control strategy design is to determine which precursor emissions to reduce, namely nitrogen oxides (NOx) and/or volatile organic compounds (VOC). Understanding the spatial and temporal variations in ozone responses to anthropogenic emissions is important. The Central Valley of California is the recipient of both locally emitted and inter-basin transported pollutants (Pun et al. 2000), which further complicates the source receptor relationship and its policy implications. Quantification of the inter-basin transport contribution and its spatial extent is currently lacking.

Chemical transport models (CTMs) integrate scientific understanding of the key physical and chemical processes for ozone formation at regional or larger scales. Past studies commonly used brute force methods conducted with CTMs to examine ozone responses to precursor pollutant changes by simulating scenarios with one-at-a-time percentage perturbations of NOx and/or VOC emissions (e.g., Dabdub et al. 1999). Brute force methods become cumbersome as the number of emission scenarios increases. In contrast, the Decoupled Direct Method (DDM) efficiently calculates the local concentration derivatives to various parameters by separately solving sensitivity equations derived from the model equations (Dunker 1981; Yang et al. 1997; Dunker et al. 2002), and has found wide applications in air quality modeling (e.g., Hakami et al. 2003, 2004; Napelenok et al. 2006). Because the accuracy of CTM simulations critically depends on the underlying input parameters, the influence of input uncertainties on ozone responses to emission-reduction strategies needs to be assessed.

In this chapter, an analysis approach with High-order DDM (HDDM) within the Community Multiscale Air Quality (CMAQ) model (Byun and Schere 2006) is developed to investigate ozone sensitivities. This chapter demonstrates this approach by exploring issues important to ozone control such as differences in control strategies in meeting the 1-h and 8-h ozone standards, changes in limiting reagent with time and space, importance of local versus upwind emission sources, and effects of emission timing. Finally, the influences of uncertainties in emission inventories, reaction rate coefficients, and boundary conditions on the effects of anthropogenic emission-reduction strategies are assessed with second-order sensitivity analyses.

[2] Reproduced in part from *Environ. Sci. Technol., 42* (10), 3683–3689, 2008.

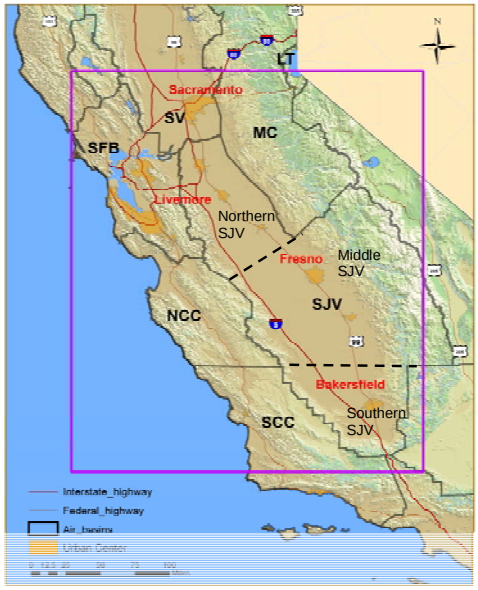
3.2 Methods

3.2.1 Study Domain and Input Data

The Central California study domain is shown in Figure 3-1, and has been described in Chapter 2. The San Joaquin Valley is further divided into northern, middle, and southern parts, as shown in the map.

Emission inputs and boundary conditions described in Chapter 2 are used. The highest anthropogenic emissions are located near urban centers and highway systems. Biogenic VOC emissions occur in vegetated regions, especially in the foothills of the Sierra Nevada and the coastal range mountains. Trainer et al. (2000) estimate that the average soil NOx for the United States is 6 percent of the anthropogenic values, and that they can be 14 percent of the anthropogenic NOx emissions in the summer. These emissions are highly variable, depending on the timing of irrigation and fertilization, and therefore are not included in the inventory. Influence of this uncertainty will be addressed later. Emissions are summarized at the whole domain and selected subregions for weekdays and weekends, respectively in Table 3-1. Influence of uncertainties in boundary conditions on ozone responses to emission reductions will be discussed later in the second-order sensitivity analysis.

Figure 3-1 Modeling Domain Indicated by the Purple Rectangle



Air basins are labeled on the map for San Francisco Bay area (SFB), Sacramento Valley (SV), Mountain Counties (MC), San Joaquin Valley (SJV), North and South Central Coast (NCC and SCC). SJV is further divided into three parts (Northern, Middle, and Southern part), as indicated by dashed lines.

Table 3-1 Emission Inputs from Selected Regions

Regions	NOx (tons/day)		AVOC (tons/day)		BVOC (tons/day)
	weekday	weekend	weekday	weekend	episode
Domain	1003	774*	1506	1625	1062
SFB	366	282	595	587	180
SJV	390	276	551	552	292
North	138	89	197	200	39
Middle	150	106	235	239	211
South	102	81	119	113	43
Rest	274	216	360	486	590

*Including fire NOx emissions (~5% of the total NOx emissions) which occurred in the South Central Coast Air Basin during weekend days.

3.2.2 CMAQ-HDDM Modeling System

Since DDM equations follow the same structure as air quality model equations, their computational details evolve with improvements in the CMAQ model. The existing implementation of HDDM (Hakami et al. 2004) was updated to work with the more recent CMAQ version 4.5, which has a number of improvements to the chemistry, Planetary Boundary Layer (PBL) modeling, and the advection scheme. New features in this version of CMAQ-HDDM include calculating sensitivities to chemical reaction rate coefficients and enabling parallel computing. These improvements are summarized and evaluated in a paper by Napelenok et al. (2007).

CMAQ-HDDM is configured to use the piecewise parabolic method for advection, multiscale horizontal diffusion and eddy vertical diffusion, and the Euler Backward Iterative (EBI) ordinary differential equation solver. Gas phase chemistry is represented using the SAPRC99 chemical mechanism (Carter 2000).

CMAQ-HDDM computes the first-order ($S_i^{(1)}$) and second-order ($S_{i,j}^{(2)}$) semi-normalized sensitivity coefficients.

$$S_i^{(1)} = P_i \frac{\partial C}{\partial p_i} = \frac{\partial C}{\partial \varepsilon_i} \tag{3-1}$$

$$S_{i,j}^{(2)} = P_i \frac{\partial}{\partial p_i}\left(P_j \frac{\partial C}{\partial p_j} \right) = \frac{\partial^2 C}{\partial \varepsilon_i \partial \varepsilon_j} \qquad (3\text{-}2)$$

where, P_i (or P_j) is an input parameter, whose perturbation p_i (or p_j) is considered in a relative sense by defining a scaling variable ε_i (or ε_j), with its nominal value (unperturbed) being 1; C is the species concentration vector; and both first- and second-order sensitivity coefficients have concentration units. A $S_i^{(1)}$ equaling α ppb implies that a ± 10 percent change in the parameter would cause a ($\pm 0.10\alpha$) ppb change in the ozone concentration while other parameters are held constant. $S_{i,j}^{(2)}$ is the sensitivity coefficient of a first-order sensitivity (Equation 3-2; it measures how a first-order sensitivity changes when another variable changes with respect to its nominal value, and can be used to explore the non-linearities in a system. $S_{i,j}^{(2)}$ equaling α ppb implies that a ± 10 percent change in the parameter *j* would cause a ($\pm 0.10\alpha$) ppb change in $S_i^{(1)}$.

3.2.3 Ozone Control Regimes

First-order ozone sensitivities to domain-wide anthropogenic NOx emissions ($\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}$) and to anthropogenic VOC (AVOC) emissions ($\frac{\partial [O_3]}{\partial \varepsilon_{E_{AVOC}}}$) are used to determine ozone-limiting reagents. Although ozone responses to emission control options are represented using a first-order approximation, they are likely to be valid for up to 25 percent emission perturbations (Vuilleumier et al. 1997), which are usually suitable for policy applications.

Three ozone control options are defined, based on the relationship between the two sensitivity coefficients.

$\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}} < 0$, VOC control,

$\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}} > \frac{\partial [O_3]}{\partial \varepsilon_{E_{AVOC}}} > 0$, NOx control,

$0 < \frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}} < \frac{\partial [O_3]}{\partial \varepsilon_{E_{AVOC}}}$, Transition.

The "VOC control" option is preferred when reducing AVOC reduces ozone concentrations and reducing NOx emissions would increase ozone concentrations (NOx disbenefit). The "NOx control" option is preferred when a percentage reduction in NOx emissions results in larger decreases in ozone concentrations than the same percentage reduction in AVOC emissions. Between these two options, the third option is in a transition regime, where reducing NOx

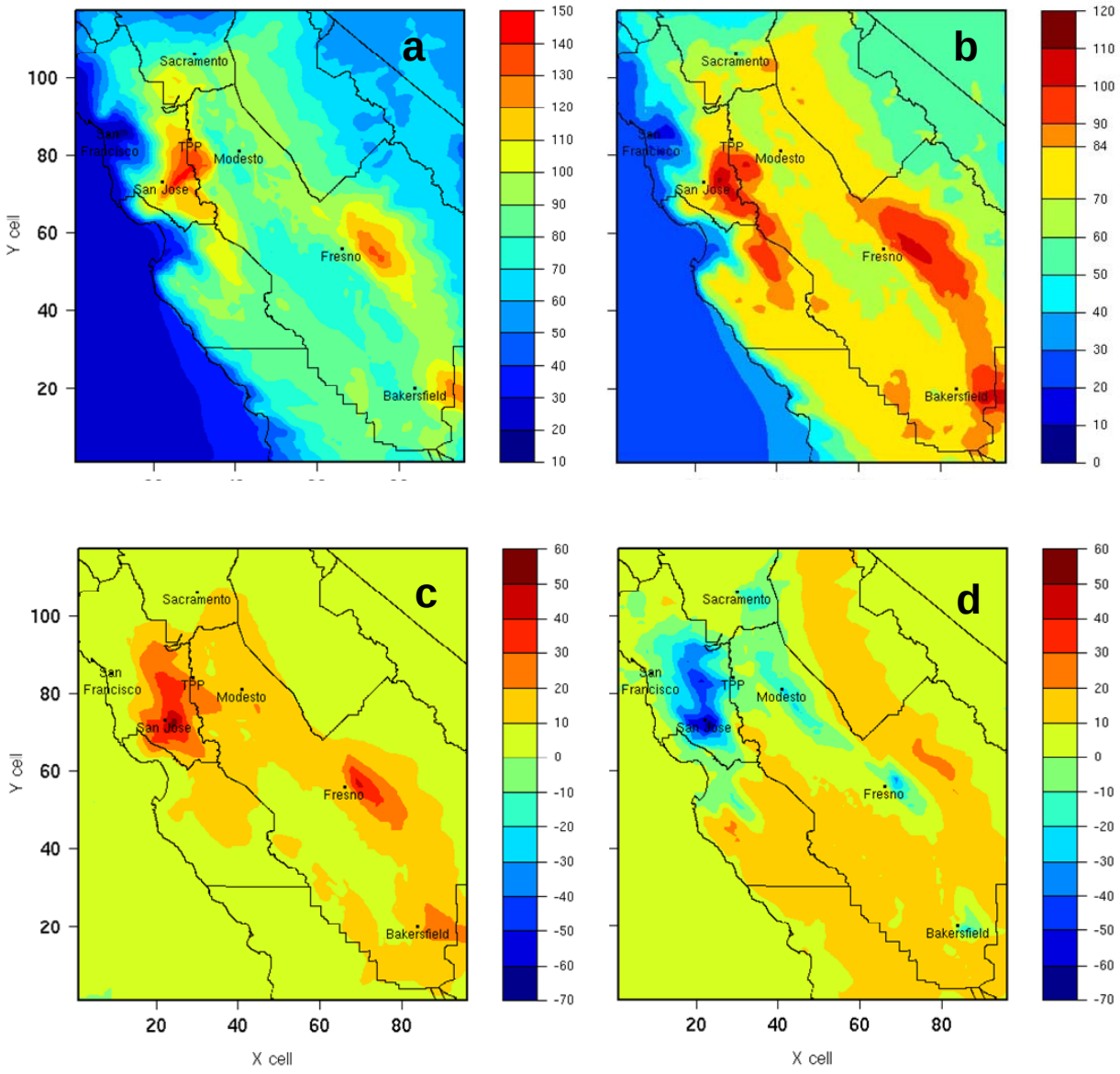
emissions can reduce ozone concentrations but is not as effective as reducing AVOC by the same percentage.

3.3 Results and Discussions

3.3.1 Simulated Ozone and First-order Sensitivity Analysis

Surface ozone concentrations and first-order ozone sensitivity coefficients to domain-wide AVOC and NOx emissions were calculated over a five-day episode (July 29–August 3), with a three-day spin-up period for both concentration and sensitivity simulations. Following the spinup period, the first two days are Saturday and Sunday, followed by three weekdays, which have similar AVOC but approximately 25 percent higher NOx emissions due to changes in human activities (Table 3-1). Temperature is increasing over the episode, leading to 18 percent higher biogenic VOC (BVOC) emission rates during the later days of the simulation. Figure 3-2 presents simulation results averaged over the weekdays, when the highest ozone levels occur. Color scales of dark orange and above in Figure 3-2ab indicate ozone exceedances. One-hour exceedances in our domain are localized in the vicinity of high NOx and AVOC sources: downwind of the San Jose area and Livermore valley, downwind of Fresno, and downwind of Bakersfield (Figure 3-2a). In contrast, modeled 8-h average ozone exceedances appear not only at locations that exceed the 1-h standard, but also include rural areas in coastal ranges and Sierra foothills, which are further downwind of high emission centers (Figure 3-2b). In these high 8-h ozone regions, $\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}}$ changes signs (Figure 3-2d) and a NOx disbenefit occurs in high NOx areas: the Bay area and major urban centers. In contrast, $\frac{\partial[O_3]}{\partial \varepsilon_{E_{AVOC}}}$ is always positive, with higher values collocated in NOx-rich areas ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} < 0$), indicating increasing AVOC reactivity with NOx supply (Figure 3-2c).

Figure 3-2: Weekday Average Peak Simulation Results (in ppb)



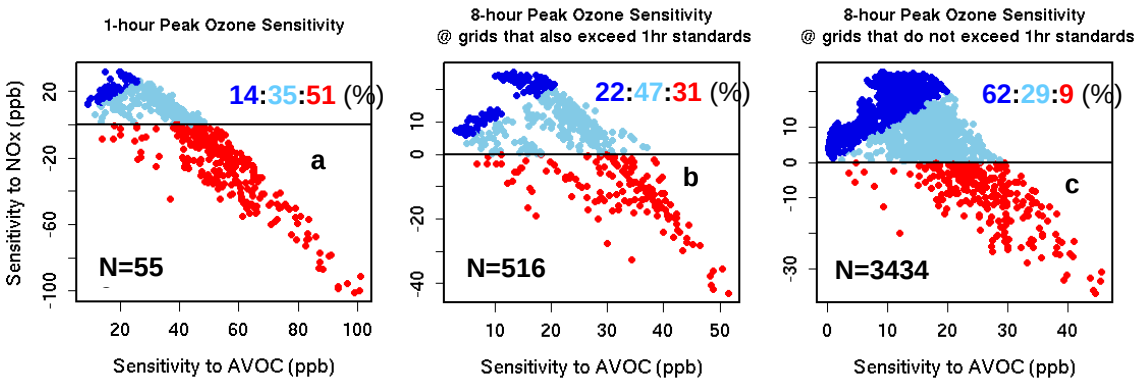
a. 1-h ozone concentrations, b. 8-h ozone concentrations, c. 8-h ozone sensitivity to domain-wide AVOC emissions, d. 8-h ozone sensitivity to domain-wide anthropogenic NOx emissions. (Color scales of dark orange and above in Figure 3-2ab indicate ozone exceedances.)

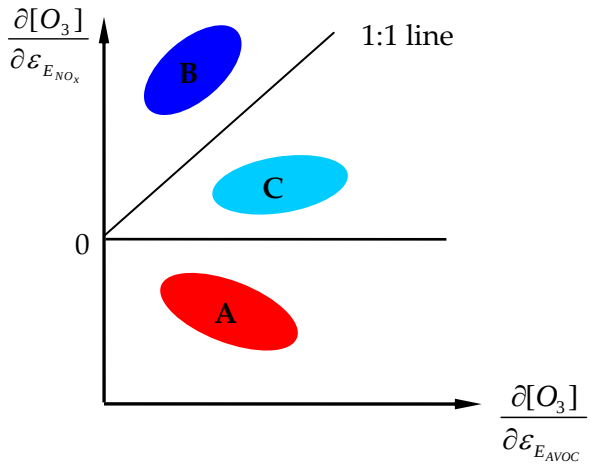
A preference for AVOC control can be readily discerned for the Bay area from Figure 3-2bcd, because high ozone tends to be located in NOx disbenefit areas. Further mapping of limiting reagents at the level of individual grid cells should be conducted for the most efficient control option in the SJV, especially for areas where both precursors have positive contributions to ozone formation.

3.3.2 Ozone Control Regimes in the SJV

As areas with high 8-h average ozone (> 84 ppb) in the SJV extend to suburban and rural areas (Figure 3-2ab), there may be quite different limiting precursors for 8-h versus 1-h peak ozone. Furthermore, even at the same location, the mean behavior captured by 8-h averages can be quite different from individual hourly values. Figure 3-3 indicates that control preferences for the 1-h and 8-h standards differ. Over 500 grid-hours of peak ozone in SJV exceed the 1-h standard, and more than 50 percent of them indicate a preference for VOC control (Figure 3-3a), while NOx control is preferred for only 14 percent of the grid hours. Approximately 4,000 grid hours of the maximum 8-h average ozone exceed the 8-h standard. The majority (86 percent) of these 8-h exceedances occurs at locations that did not exceed the 1-h standard; more than 60 percent of the grid-hours indicate a preference for NOx control; whereas, only 9 percent prefer VOC control (Figure 3-3c). Despite spatial variations, NOx control is overall more beneficial for reducing 8-h peak ozone in the SJV, in contrast to VOC control, which works better for attaining the prior 1-h standard.

Figure 3-3: Ozone Control Options for Peak Ozone Exceedances During the Five-Day Episode





Definition of Ozone control options

A: $\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} < 0$, VOC control,

B: $\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} > \frac{\partial[O_3]}{\partial \varepsilon_{E_{AVOC}}} > 0$, NO_x control,

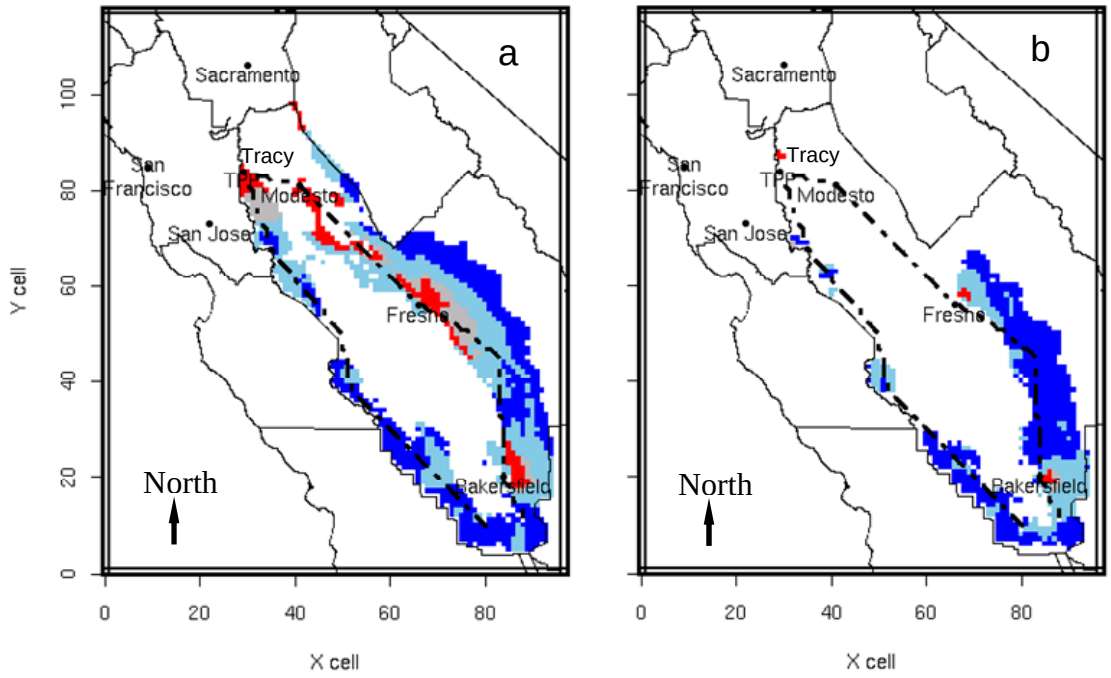
C: $0 < \frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} < \frac{\partial[O_3]}{\partial \varepsilon_{E_{AVOC}}}$, Transition.

Since preferred ozone control option regimes change in both time and space as a function of emissions and meteorology (Blanchard and Stoeckenius 2001; Baertsch-Ritter et al. 2004; Couach et al. 2004), this research determined how ozone control options vary throughout the SJV and for different days. Daily ozone control options are determined in the SJV at individual grid cells where 8-h peak ozone exceeds the 8-h standard, then the results are mapped in Figure 3-4 according to the daily control options and whether they change across different days. Grid cells where the sign of ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}}$) changes from day to day are colored gray. Weekends and weekdays are presented separately, for they differ in both emissions and meteorology during the simulation period. High ozone levels (> 84 ppb) are found along Highway 99 and Interstate 5 (marked in Figure 3-1) and areas downwind on both weekdays and weekends. The VOC control option is preferred for metropolitan areas in the SJV and places near Bay area and Sacramento emission sources for the weekdays (Figure 3-4a). NOx control is preferred in the Sierra foothills and coastal ranges. Between urban and rural settings, where average $\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}}$ is about to change signs (Figure 3-2c), ozone control options may change day by day (the grey areas in Figure 3-4), in that NOx control can be beneficial ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} > 0$) or detrimental (i.e., $\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} < 0$) at the same location on different days. Since weekday emissions are modeled to be almost identical, these changes are most likely caused by day-to-day differences in meteorology. Temperatures on the weekends are 3 to 5 degrees (K) cooler, resulting in less extensive ozone exceedances seen in Figure 3-4b. Compared to the weekday sensitivity map, ozone control on weekends in this episode would be better served by NOx control, including a number of rural areas that were in "NOx transition" on weekdays. This is partly due to greater ozone sensitivities to NOx emissions when NOx levels are lower on weekends, which is referred to as a *concave daytime ozone response* to NOx emissions (Cohan et al. 2005).

Reynolds et al. (2003) utilized analyses of ambient measured ozone precursor data with model results conducted by the California Air Resources Board using the SAQM photochemical model. That study considered emission change scenarios with 1990, 1999, and 2005 as the base case for the same domain, but for a different ozone episode. In contrast to the formal sensitivity results presented here, they used the photochemical model results to construct EKMA-like ozone isopleths to study precursor and ozone relationships in various sub-domains, and performed additional modeling studies using Process Analysis to investigate the NOx chemistry. They found that substantial NOx reductions were required to achieve compliance with the 8-h standard, but they did not explore what might happen spatially in the entire domain and how control options change from day to day.

This study found that outside of the larger urban areas, NOx emission reductions are necessary to comply with the 8-h ozone standard, in agreement with Reynolds et al., and the magnitudes of our largest ozone-to-NOx sensitivity coefficients indicate that large reductions in NOx would be required to achieve compliance with the standard, also in accord with Reynolds et al.

Figure 3-4: Mapping of Ozone Sensitivity Options at 8-h Peak Ozone Exceedances in the SJV



a. weekdays, b: weekends. Locations always in "VOC control" are coded in red; NOx control always beneficial ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} > 0$) in light blue, among which, the locations that are always in "NOx control" option ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}} > \frac{\partial[O_3]}{\partial \varepsilon_{E_{AVOC}}}$) are further coded in blue; and finally, a mixed regime ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}}$ changes signs) in grey. The non-exceedance areas in the SJV are marked in white.

3.3.3 Local Versus Upwind Source Contributions

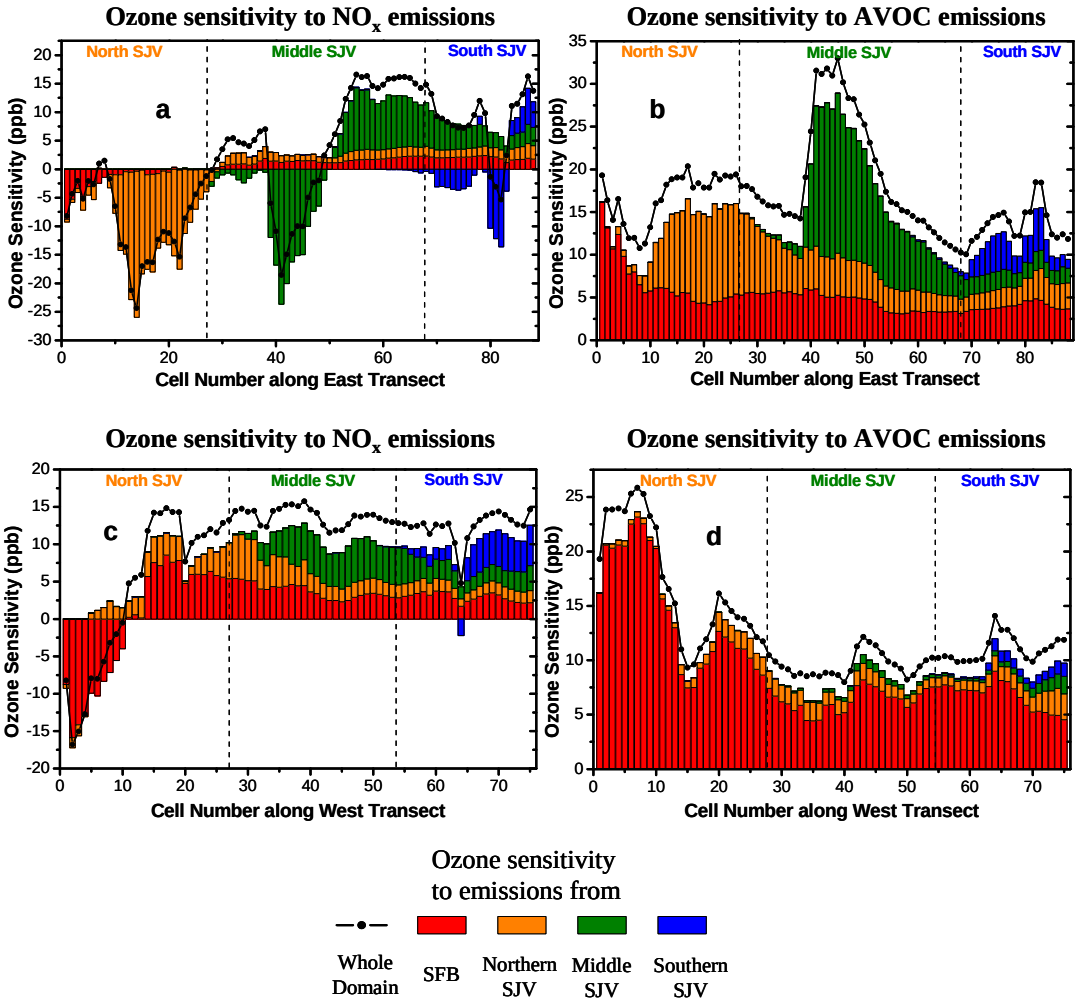
Typically, emissions over the whole modeling domain are perturbed for calculating ozone sensitivities, so that all possible anthropogenic emissions that may influence air quality are included. Consequently, the ozone control option derived from these sensitivities is a "regional decision." In reality, air quality is usually managed at air basin scales. The first-order ozone sensitivity to domain-wide emissions for any location is equal to the sum of ozone sensitivities to emissions associated with each air basin, or even further divided subregions. These various sensitivities may vary greatly in magnitude and sign. Understanding contributions from each subregion is particularly important for the SJV because it is downwind of large emission sources, and their role in SJV air pollution is not understood. It is not clear whether the "regional decision" should be applied to SJV alone or whether upwind air basins should be included in order to achieve effective ozone control. Another goal of the study was to develop an analysis approach that could be used to determine the relative importance of intra- and inter-basin transport, which could have implications to power plant siting, as well as emissions

trading between power generation sources. A scoping study on the potential role of credit trading and using this approach to better understand impacts of such approaches is discussed in a companion report.[3]

Ozone sensitivity to Bay area emissions is used to illustrate the importance of inter-basin contributions. Ozone sensitivities to emissions from the northern, middle, and southern SJV (Figure 3-1) are calculated so that local and intra-basin contributions can be further distinguished. For example, at locations in the southern SJV, ozone sensitivities to southern SJV emissions represent the local contribution, while ozone sensitivities to emissions from other parts of the SJV represent intra-basin contributions. Local, intra-basin, and inter-basin source contributions to the peak 8-h ozone levels at grid cells along two transects (black dash-dotted lines in Figure 3-4) are averaged over the weekdays and presented in Figure 3-5. The western transect is along the I-5 transportation corridor, which runs along rural areas and passes through major locations that are out of compliance with the 8-h standard. The east transect is along Highway 99, which runs through major SJV metropolitan areas and their downwind areas, where 8-h ozone exceedances also occur. Both transects start from Tracy station (indicated in Figure 3-4) in the northern SJV and move downwind away from the Bay area toward the southern SJV. Note that the subregional contributions considered here do not add up to ozone sensitivity to domain-wide emissions (the black dotted line, hereinafter called "regional contribution"), because emissions from other air basins also play a small role. A semi-normalized ozone sensitivity of less than 5 ppb is considered to be insignificant, because for a typical 20 percent emission reduction, a 5 ppb sensitivity would translate to 1 ppb ozone difference.

[3] The parallel report conducted as part of the project focuses is called *Impact of Short–Term, Interbasin, and Interpollutant Credit Trading on Air Quality and Credit Prices,* and it can be found on the California Energy Commission's website.

Figure 3-5: Weekday Average 8-h Peak Ozone Sensitivities and Decomposition of Contributions from Subregions



a. ozone sensitivity to NOx emissions along the east transect; b. ozone sensitivity to AVOC emissions along the east transect; c. ozone sensitivity to NOx emissions along the west transect; d. ozone sensitivity to AVOC emissions along the west transect. The transects run from north to south along the San Joaquin Valley, and are shown as dashed lines in Figure 3-4.

Along the eastern transect, ozone responses (sign and magnitude) to changes in domain-wide emissions are largely determined by local contributions, with only a few exceptions (Figure 3-5ab): emissions from the Bay area determine ozone sensitivities at locations near Tracy in the northern SJV; AVOC contributions from the Bay area and other parts of the SJV account for more than 50 percent of the ozone responses in the southern SJV, as it is downwind of all the other emission sources. Sign difference between regional and local contributions occurs at a few places between the major emission sources (Figure 3-5a). For example, upon entering the southern SJV, positive contributions of NOx emissions from the middle SJV and further upwind areas dominate the local NOx disbenefit, so that ozone sensitivity to domain-wide NOx

emissions is positive. At places like this, when regional NOx control is applied locally, it must be accompanied by NOx controls in upwind areas to achieve ozone reductions. The general trend for sensitivity to NOx is that local emissions exhibit a NOx disbenefit near emission sources, but contribute to ozone formation as the formed odd oxygen (NO_2 and O_3) is transported downwind. Inter-basin contributions due to NOx emissions from the Bay area are only important for the northern SJV. In contrast, positive and significant $\frac{\partial[O_3]}{\partial \varepsilon_{E_{SFBAVOC}}}$ indicate that Bay area AVOC emissions influence ozone formation in the northern and the middle SJV.

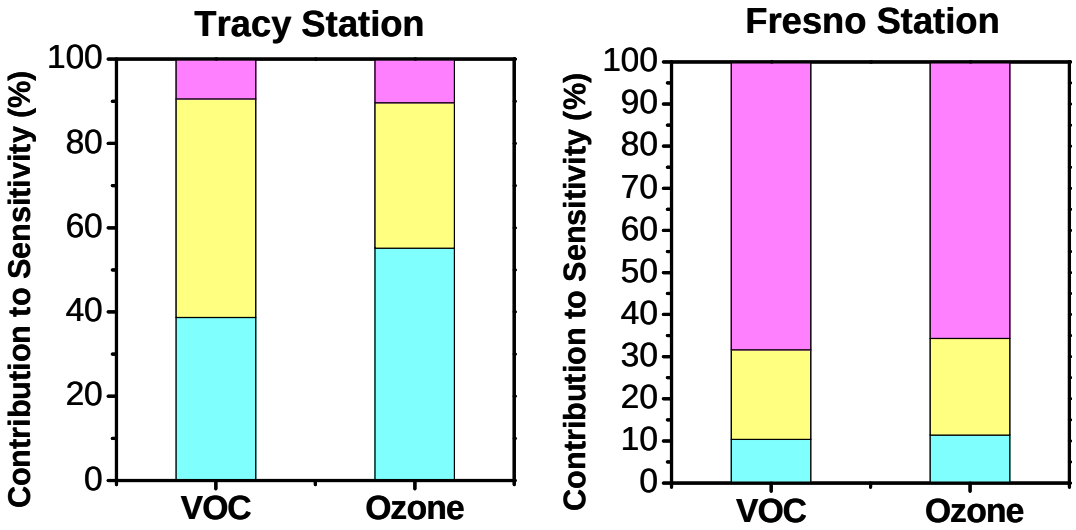
Along the western transect, the influence of Bay area emissions reaches further downwind for the ozone episode considered here (Figure 3-5cd). Bay area NOx emissions largely determine the sign and magnitude of regional contributions to ozone responses in the northern SJV, while in the middle and southern SJV, local and intra-basin contributions are more important. A positive and significant influence of Bay area AVOC emissions dominates ozone formation along the western transect. Therefore, the VOC control option preferred in the Bay area not only benefits local air quality, but also reduces downwind ozone significantly. Along both transects, NOx emitted in the Bay area has a shorter lifetime for influencing SJV ozone formation than Bay area AVOC emissions.

Bay area AVOC emissions have been shown to contribute to downwind ozone in the northern and middle SJV along the east transect, which passes a number of VOC-limited urban locations. Bay area AVOC emissions can contribute to the SJV ozone concentrations in two ways: (1) by being transported to the SJV, thus increasing the SJV local precursor budget, and (2) by forming ozone en route, thereby directly increasing SJV ozone levels by transport. In both cases, emission timing is important for inter-basin transport of pollutants because photochemical production of ozone only occurs during the daytime hours.

Bay area AVOC emissions were perturbed in three different time intervals: morning (5 a.m. to noon), afternoon (noon to 7 p.m.), and evening (7 p.m. to next day 5 a.m.). Their relative contributions to the increase of the VOC and ozone concentrations due to Bay area AVOC emissions ($\frac{\partial[VOC]}{\partial \varepsilon_{E_{SFBAVOC}}}$ and $\frac{\partial[O_3]}{\partial \varepsilon_{E_{SFBAVOC}}}$) were calculated and averaged during the peak 8-h ozone at Tracy in the northern SJV and Fresno in the middle SJV (Figure 3-6). Although the VOC concentrations in the northern SJV are most sensitive to Bay area afternoon emissions (~55 percent of $\frac{\partial[VOC]}{\partial \varepsilon_{E_{SFBAVOC}}}$), ozone levels are most sensitive to Bay area morning emissions (~50 percent of $\frac{\partial[O_3]}{\partial \varepsilon_{E_{SFBAVOC}}}$). The most reactive AVOC emissions from the Bay area are consumed before reaching Tracy, but the resulting odd oxygen (O_3+NO_2) that is formed contributes to ozone levels by transport. In contrast to Tracy, Fresno is further downwind from the Bay area, and at Fresno, both VOC and ozone are most sensitive to Bay area evening emissions (70 percent) from the previous day. Bay Area evening VOC emissions are transported further

downwind in an air mass that is not photochemically active. Air masses containing Bay area evening VOC emissions pass by Tracy and arrive at Fresno when the sun rises, so they can actively participate in daytime ozone production. Evening emissions contribute about the same percentage increase to VOC and ozone concentrations, which indicates that they increase ozone concentrations in the middle SJV by increasing its VOC budget (Figure 3-6).

Figure 3-6: Contributions to $\frac{\partial[O_3]}{\partial \varepsilon_{E_{SFBAVOC}}}$ and $\frac{\partial[VOC]}{\partial \varepsilon_{E_{SFBAVOC}}}$



Contributions from SFB AVOC emitted in the morning (cyan), in the afternoon (yellow), and in the evening (magenta). Results are shown at two locations, each covering 20 km by 20 km.

Pun et al. (2000) proposed a conceptual model describing inter- and intra-basin transport of pollutants that leads to high ozone concentrations in the SJV. They concluded that transport of ozone and its precursors from the Bay area mainly affects the northern part of the SJV, and local emissions and chemical transport within the SJV are responsible for the regional and urban-scale distribution of ozone. We developed a quantitative method to assess the importance of inter- and intra- basin emission sources and their timing on ozone formation and its accumulation in high ozone areas in both urban and rural settings. In the episode considered here, differentiated contributions from Bay area NO_x and AVOC emissions to the downwind SJV air basin were found. Anthropogenic VOC emissions, especially those emitted at night, influence ozone levels in the SJV further downwind. In particular, along the western transect, the influence of Bay area AVOC emissions dominates local AVOC sources within the SJV.

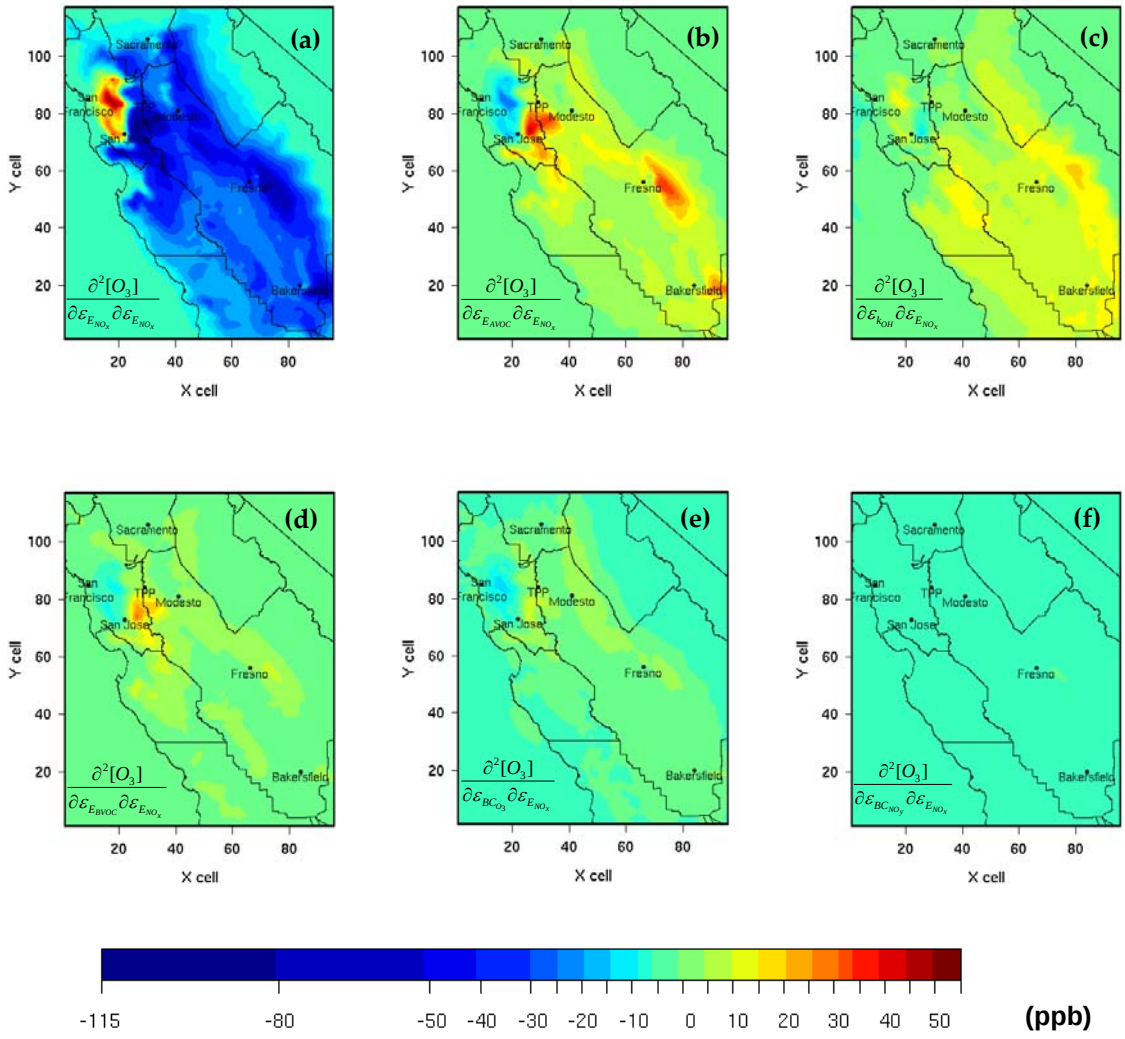
Two things should be noted for this analysis. First, the relative and absolute inter-basin contributions reported here are only representative of this particular ozone episode. As described in Steiner et al. (2006), during this episode, a high-pressure system over the Great Basin created an offshore pressure gradient that reduced westerly flows from the Bay area into the SJV. Consequently, Bay area emissions are less connected to the eastern part of the SJV, and more to the western part (see flow patterns in Steiner et al. 2006). Inter-basin contributions to the eastern transect in the SJV could increase during other ozone episodes, when westerly winds are stronger. Second, the importance of inter-basin and intra-basin emissions is compared based on source contributions (same percentage perturbations in emission sources), into which the regional contribution ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{AVOC}}}$ or $\frac{\partial[O_3]}{\partial \varepsilon_{E_{NOx}}}$) can be decomposed. Since Bay area emissions are about three times those from each of three parts of the SJV (Table 3-1), if ozone sensitivities to per-tonnage emissions are considered, the importance of Bay area emissions on the SJV ozone levels would be less pronounced in the middle and southern SJV.

3.3.4 Influence of Uncertainties in Other Parameters

Findings with first-order sensitivity coefficients are applicable to small emission reductions for a given set of model inputs and parameters, which can be uncertain and vary over time. Second-order sensitivities measure the sensitivity of first-order sensitivities to other input variable changes, and are used here to quantify how ozone responses to NOx emissions are influenced by other model uncertainties (Cohan et al. 2005). ($\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}}$) changes with uncertainties in other model inputs are considered because our results indicate that NOx control is preferred for abating 8-h peak ozone in the SJV. Also, Hakami et al. (2004) found that NOx emissions in Central California produced the most non-linear ozone response (the largest magnitude of higher order sensitivity coefficients). Six model input parameters were selected for the second order sensitivity of $\frac{\partial[O_3]}{\partial \varepsilon_{E_{NO_x}}}$: NOx emissions, AVOC emissions, rate coefficient of the $OH+NO_2$ termination reaction, BVOC emissions, O_3 inflow boundary conditions, and NO_y[4] inflow boundary conditions. Large second-order sensitivities usually appear when both first-order sensitivities are large (Vuilleumier et al. 1997; Brown and Revzan 2005). Hence, the selection of the second variable is based on relatively large first-order sensitivity as well as high uncertainty. All second-order sensitivity coefficients are semi-normalized to facilitate comparison.

[4] NO_y is total reactive nitrogen.

Figure 3-7: Weekday Average Second-order Sensitivities of 8-h Peak Ozone



These are sensitivities of $\frac{\partial[O_3]}{\partial\varepsilon_{E_{NO_x}}}$ to (a) NOx emissions, (b) AVOC emissions, (c) OH+NO_2 rate coefficient, (d) BVOC emissions, (e) O_3 inflow boundary condition, and (f) NOy inflow boundary condition.

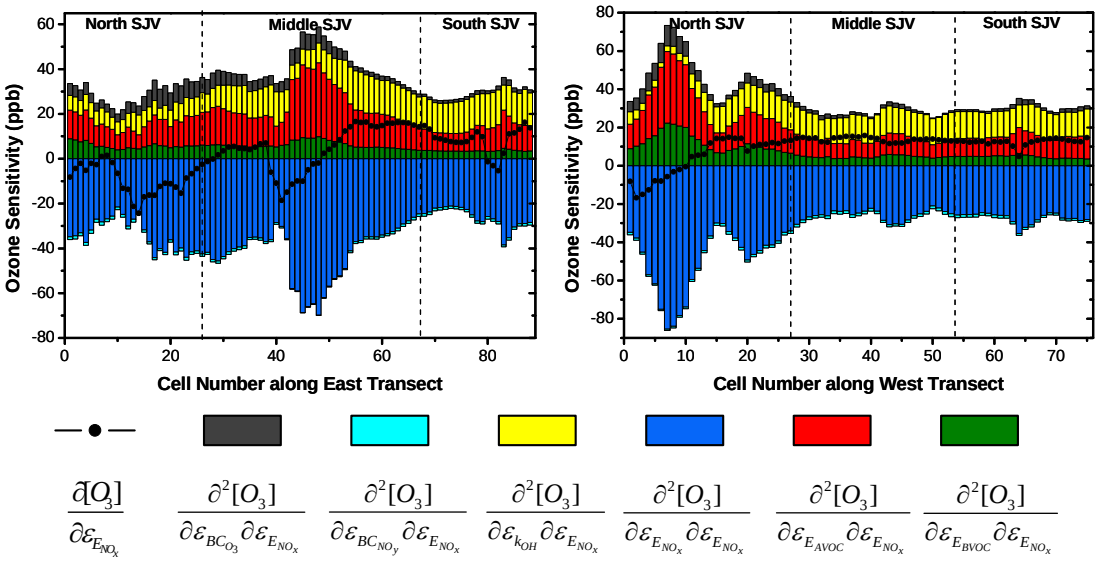
Weekday second-order sensitivity maps are shown in Figure 3-7; the sensitivities are averaged at the peak 8-h ozone hours, and are shown in order of decreasing magnitude. $\frac{\partial}{\partial\varepsilon_{E_{NO_x}}}\left(\frac{\partial[O_3]}{\partial\varepsilon_{E_{NO_x}}}\right)$ has the largest magnitude, and is negative over most of the domain, except part of the Bay area with low ozone concentrations (Figure 3-7a and Figure 3-2ab). There are abrupt changes in the sign

of second-order sensitivities in Figure 3-7a,b, and d. This is due to a transition from a region where NOx titrates ozone to one where more photochemical ozone production occurs.

As mentioned earlier, biogenic NOx emissions from the soil can be 14 percent of the anthropogenic NOx emissions in the summer, but they are not included in the emission inventory. Therefore, according to $\frac{\partial}{\partial \varepsilon_{E_{NO_x}}}\left(\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}\right)$, excluding these emissions would produce a positive bias of 2~6 ppb in the ozone response in the SJV to NOx emissions, which leads to an over-estimation of the impact of a specific NOx control option. $\frac{\partial}{\partial \varepsilon_{E_{AVOC}}}\left(\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}\right)$ is mostly positive, except in part of the Bay area with low ozone concentrations, and it is large where $\frac{\partial}{\partial \varepsilon_{E_{NO_x}}}\left(\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}\right)$ is large but of the opposite sign. The magnitudes of $\frac{\partial}{\partial \varepsilon_{E_{NO_x}}}\left(\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}\right)$ and $\frac{\partial}{\partial \varepsilon_{E_{AVOC}}}\left(\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}\right)$ tend to reach their maxima at places where the first-order sensitivity ($\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}$) is about to change signs (Figure 3-8), corresponding to the grey areas between urban and rural settings in Figure 3-4a. Therefore, not only are control options in these grey areas more sensitive to meteorological variations than typical urban or rural areas, as found earlier, but also their accuracy is more susceptible to emission uncertainties. The OH+NO_2 chain terminating reaction, which removes HOx and NOx from the air mass, has been found to be one of the most important reactions in determining ozone production (Martien and Harley 2006). According to a recent Jet Propulsion Laboratory (JPL) evaluation (Sander et al. 2006), the uncertainty factor associated with the reaction rate coefficient is large and varies with temperature ($\pm$30 percent at 298 K). This rate coefficient is the third most influential variable for ozone responses to NOx emissions, and the second-order sensitivity to it is mostly positive throughout the SJV, where increasing it would make ozone sensitivity to the remaining NOx more positive. If it were revealed that this rate coefficient was 30 percent too low, eliminating the error would result in increasing the sensitivity of ozone to NOx. At locations downwind of Bakersfield (see Figure 3-8), ozone would be more responsive to NOx emissions by about 6 ppb, which would increase the effectiveness of the NOx control option. Biogenic VOC emissions are the largest VOC source in the domain; however, their influence on $\frac{\partial [O_3]}{\partial \varepsilon_{E_{NO_x}}}$ is much less than that of AVOC emissions. Since cross derivatives of two pollutant emission sources increase with their proximity to each other (Cohan et al. 2005), and BVOC is generally not collocated with high NOx emissions, its influence is greatly reduced. Nevertheless, doubling BVOC emissions inputs would increase ozone sensitivities to NOx emissions by 10 ppb in certain places (e.g., downwind of Fresno, see Figure 3-8), and thus increase the response to NOx control. Our assumption of constant boundary conditions introduces uncertainties to the modeling system as boundary conditions vary by season and are influenced by inter-continental transport of pollution (Nowak 2004). Uncertainties in boundary ozone could change ozone response to NOx

emissions by about 5 ppb; whereas, the influence from specification of boundary NOy concentrations is negligible, assuming a clean marine air mass as inflow.

Figure 3-8: Weekday Average of Second-Order Sensitivity Coefficients of 8-h Peak Ozone Along East and West Transects



CHAPTER 4:
Seasonal Versus Episodic Performance Evaluation for the CCOS Domain[5]

4.1 Introduction

The Community Multiscale Air Quality Model (CMAQ) (Byun and Schere 2006) is widely used to study air pollutant formation and transport. Model performance for specific geographic regions is mostly evaluated for short episodes (e.g., Mebust et al. 2003; Zhang et al. 2005, 2006), and studies often focus on multi-state regions in the eastern United States. While longer-term simulations are desirable, fewer studies are conducted mainly due to lack of adequate supporting data for model input preparation and performance evaluation, as well as high computational cost. Hogrefe et al. (2001, 2004a, 2006) evaluated CMAQ's ability to capture summer ozone variability observed in present-day climate over the eastern United States, focusing on the temporal features of the distributions. Eder et al. evaluated a forecast model with pair-wise comparisons between simulated ozone and extensive ground measurements over a summer season for the eastern U.S. domain (2006) and the whole continental United States (2009). Similar pair-wise comparison was conducted by Tong and Mauzerall (2006) for the whole continental United States to evaluate predicted ozone spatial variability over one summer month. These longer-term simulations and performance evaluations generally relied on routine emission inventories and measurement networks; ozone precursor predictions were often not addressed. The coarser horizontal resolution (e.g., 36 km used in Tong and Mauzerall 2006 and 12 km used in Eder et al. 2009) may affect the model's ability to represent variable terrain and land use in the mountainous west. The Central California region has had long-term and serious ozone problems (Jin et al. 2008) and understanding CMAQ performance here is critical for model application.

The meteorology inputs are an integral part of photochemical modeling. Uncertainties in these inputs affect not only the simulated ozone levels, but also ozone responses to emission changes (Biswet al. 2001; Jin et al. 2008). Ideally, meteorology simulations can be evaluated in conjunction with chemical predictions to interpret episode-to-episode and region-to-region differences in model performance. Such efforts can only be supported when sufficient data are available and are commonly not pursued together with chemical evaluations.

In this chapter, we extend our modeling to the full Central California Ozone Study (CCOS) domain, which includes the whole Sacramento Valley. We evaluate seasonal and episodic model performance with MM5-CMAQ version 4.5 applied to the entire summer of 2000 in Central California, where significant spatial and temporal variations in air quality are observed. A fine grid spacing (4 km) is used to account for the complex terrain of the study domain. A rich aerometric database collected during the summer 2000 CCOS, including ozone and its precursor species, is used to benchmark model results. Emission inputs are developed for the CCOS

[5] Reproduced in part from Jin et al. 2010.

period that manifest variability occurring on diurnal, weekly, and seasonal time scales. Furthermore, a network of 25 wind profilers was deployed to collect information on the winds and temperature structure in the lower troposphere in the Central Valley, to improve the understanding of meteorology in this region. These data are used to assess the impact of uncertainties in meteorological fields on chemical outputs. Both ozone and its precursor predictions are evaluated in conjunction with the assessment of MM5 modeled meteorological fields. In addition to assessing temporal and spatial behaviors of ozone predictions, we describe the similarity and differences in seasonal and episodic performance of the meteorological and chemical simulations. Evaluation metrics are presented for the full summer season and for four short intensive operating periods (IOPs) during CCOS that represent distinct meteorological conditions where high ozone concentrations were observed.

4.2 Data and Methods

4.2.1 Study Domain

The CCOS domain shown in Figure 4-1 extends from approximately 34°N to 41°N and 116°W to 124°W, is modeled using a 185 × 185 grid with a horizontal resolution of 4 km. Vertically, the domain is divided into 27 layers from the surface to 100 millibar (about 17 km); the near-surface layers are about 20 m thick. Air basins in the study domain include the SJV, SFB, SV, Mountain Counties (MC), and several others, as shown in Figure 4-1.

The inner domain (SARMAP domain 96 × 117 grid cells) includes the entire SJV and SFB, as well as the Sacramento metropolitan area that contains most of the emissions from the SV. The smaller domain has been used for more detailed mechanistic and diagnostic evaluations conducted over shorter time periods, as described in Chapters 2 and 3.

4.2.2. Emission Inputs

Emission inputs have been considered as a major source of uncertainty in air quality modeling (Russell and Dennis 2000). For long-term modeling, it is important to assign adequate variability to the emission inputs, to introduce appropriate variability in model outputs. Emission inventory files were provided by ARB, developed for the CCOS period in five source categories: area sources, motor vehicle, point sources, biogenic, and forest fires. The anthropogenic emissions are estimated with data for a one-week intensive operating period. Off-road mobile sources were inventoried and included with area sources. Emission inputs were first checked for quality assurance, and a description of this is provided in Appendix A.

Anthropogenic emissions (area, mobile, and point sources) are specified with day-of-week differences (Figure 4-2), and the same weekly cycle was repeated for the whole summer. Area source NOx emissions are generally lower (~16 percent domain wide) on weekends, but VOC emissions are estimated to be higher (~19 percent domain wide) on weekends because of increased recreational activities (e.g., boating). For motor vehicle emissions, relative to weekdays, there are large reductions in diesel truck traffic and associated NOx emissions on weekends. On average, weekend mobile NOx emissions are about 25 percent less than those on weekdays, while reductions in on-road mobile VOC emissions are less than 10 percent. Individual point sources are described by location, stack parameters (diameter, height, exit

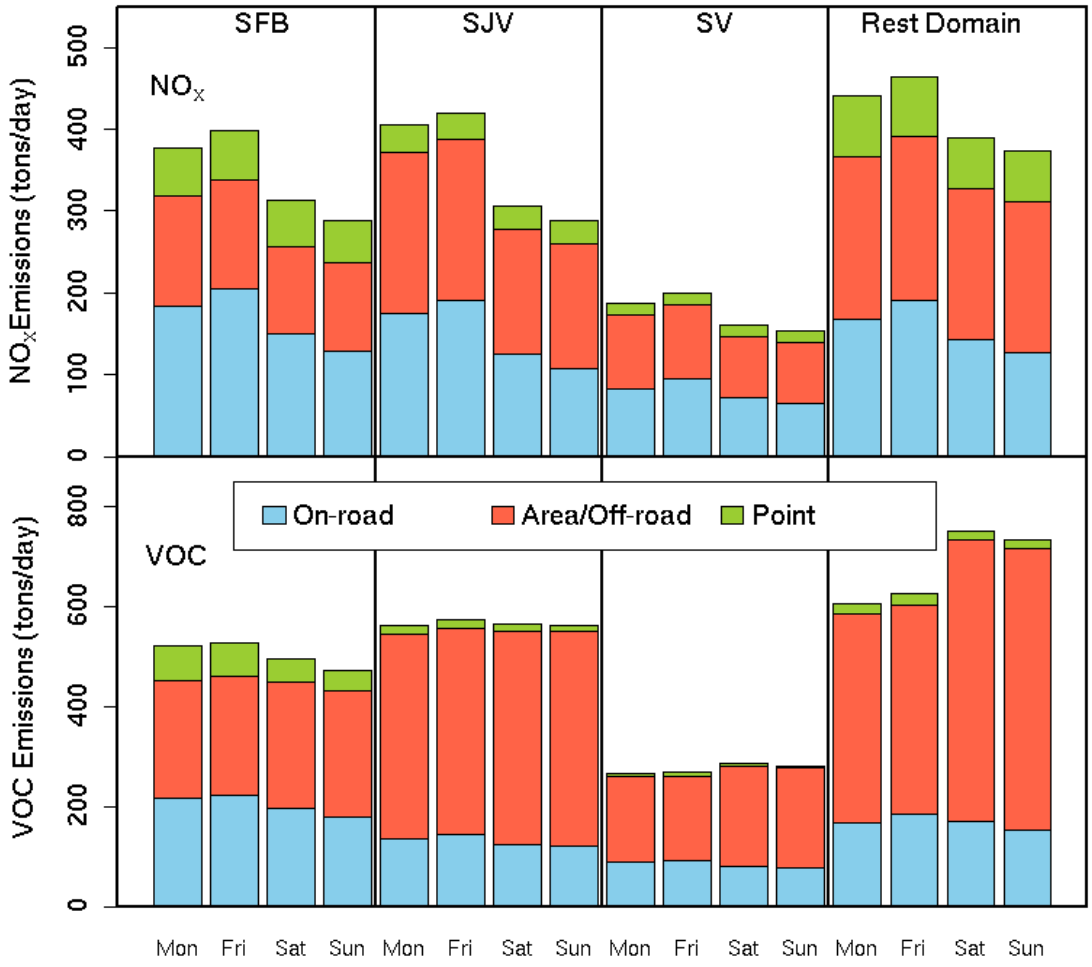
temperature, exit velocity), and hourly emissions of pollutants. Plume rise is estimated for each point source using the emission preprocessor SMOKE version 2.1 (Coats 1995) driven by day-specific MM5 meteorology.

Hourly biogenic emissions (mainly isoprene and terpenes) were estimated for each individual day of the summer season. Emissions were estimated using the BEIGIS modeling system (Scott and Benjamin 2003) by applying date-specific temperatures and sunlight intensity to seasonally adjusted leaf-cover estimates. Daily emissions in metric tons are shown in Figure 4-3 for isoprene and terpenes. Isoprene emissions occur only in the presence of light, and increase with solar radiation until a saturation point is reached. The emissions increase exponentially with temperature up to approximately 35°C–40°C, after which they decrease. Monoterpene emissions increase exponentially with temperature and are not dependent on light. These light and temperature effects are represented in the model using parameterizations (Guenther et al. 1995). Total biogenic emissions vary on a daily basis from 500 tons to as much as 3,500 tons, depending on meteorology. Biogenic emissions are especially prevalent in the foothills of the Sierra Nevada and the coastal mountain ranges, and are generally not collocated with large anthropogenic emission fluxes.

Figure 4-1: SARMAP and CCOS Domain

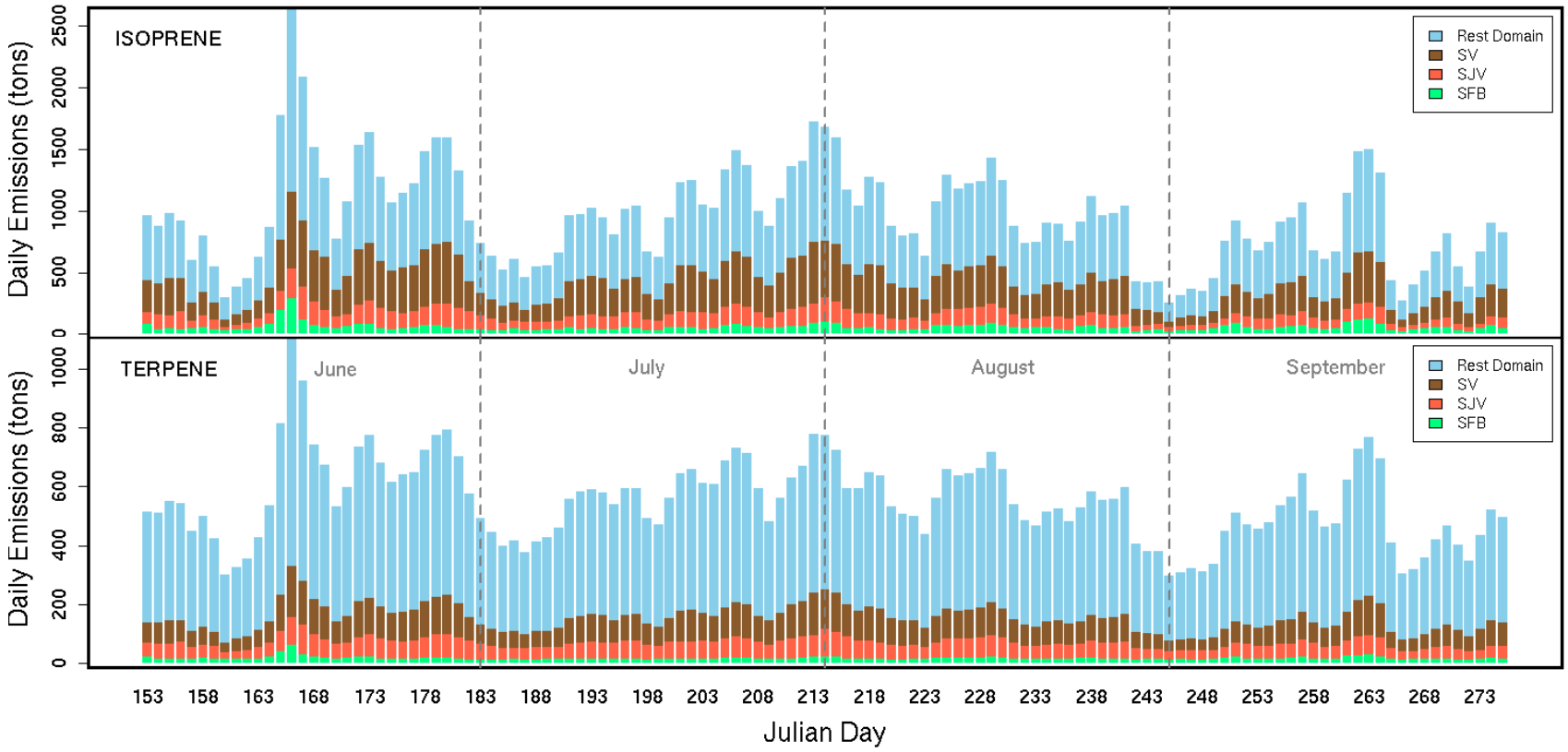
CCOS domain shown in the outer red rectangle, a grid of 185 by 185 with each cell having a resolution of 4 km by 4 km. The inner purple rectangle shows the previous SARMAP domain, a 96 by 117 grid. Major air basins labeled on the map are: San Joaquin Valley (SJV), Sacramento Valley (SV), San Francisco Bay area (SFB), Mountain Counties (MC), and North Central and South Central Coasts (NCC and SCC).

Figure 4-2: Daily Total Emissions in Summer 2000



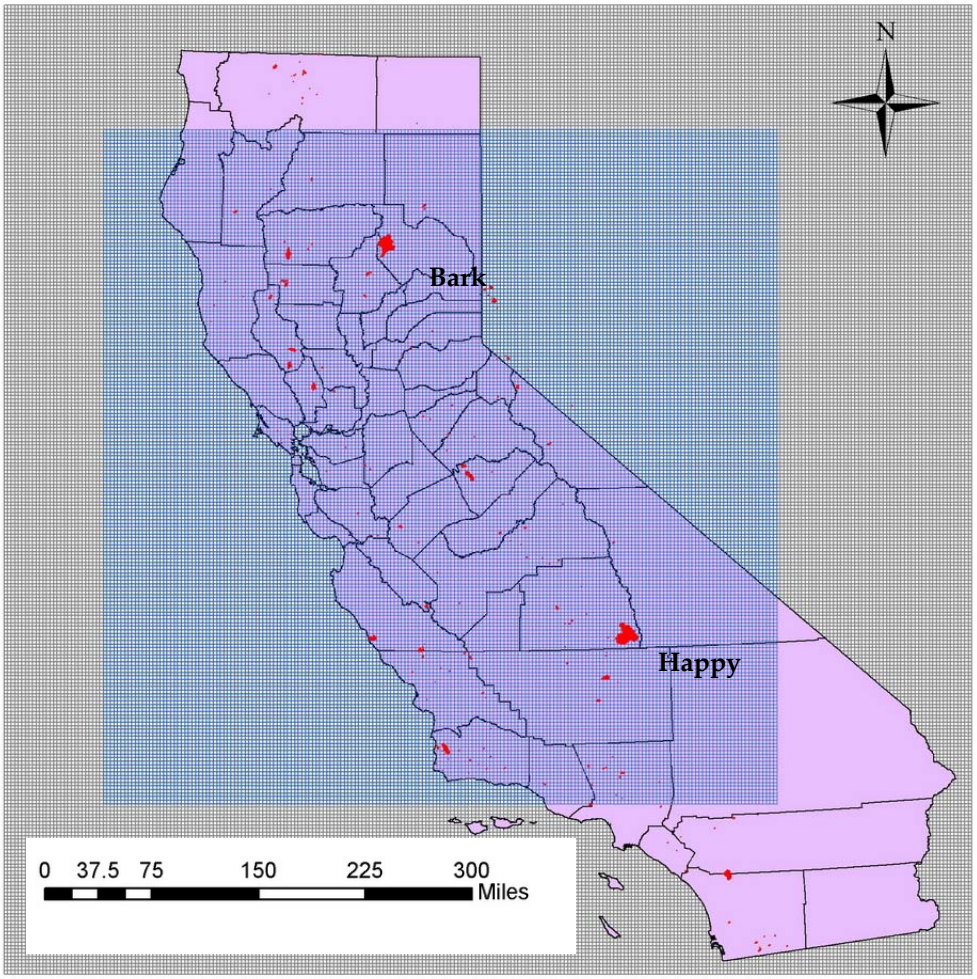
Monday emissions also were applied to other weekdays, Tuesday through Thursday.

Figure 4-3: Daily Total Isoprene and Terpene Emissions in Summer 2000



Forest fire emission estimates were made by ARB for the whole summer with the Emission Estimation System (EES) model (http://www.arb.ca.gov/ei/see/see.htm). Information includes daily emissions in metric tons for nine species (PM_{10}, $PM_{2.5}$, CO, CH_4, NMHC, NH_3, N_2O, NO_x, and SO_2)[6] and associated fire name, date, acres burnt, and location of the fire. Figure 4-4 shows the locations of the fires and the study domain for summer 2000.

Figure 4-4: Location of California Forest Fires During Summer 2000

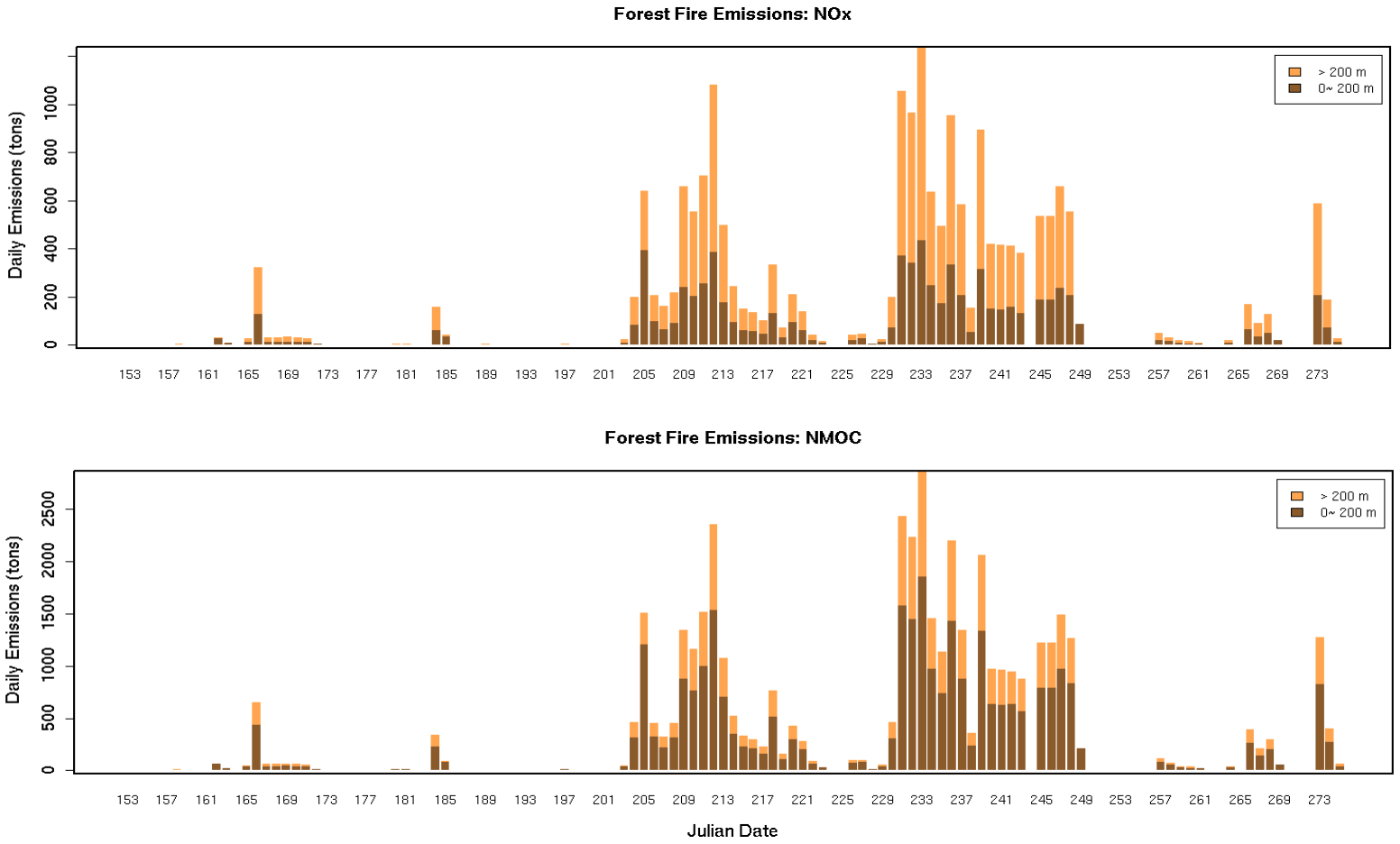


Red dots show the fire locations. CCOS domain grids are shown in blue. The state grids are shown in grey.

[6] PM_{10} = particulate matter ≤ 10 micrometers, $PM_{2.5}$ = particulate matter ≤ 2.5 micrometers, CO = carbon monoxide, CH_4 = methane, NMHC = non-methane hydrocarbons, NH_3 = ammonia, N_2O = nitrous oxide, NO_x = nitric oxide and nitrogen dioxide, and SO_2 = sulfur dioxide.

These daily total emissions were speciated into SAPRC 99 species and allocated hourly according to the diurnal profiles that depend on species types. Nitrogen oxide follows a diurnal profile that has a relatively sharp rise in the middle of the day (FEJF 1996), except on smoldering days. Carbon monoxide, SOx, and NMHC follow a flatter profile throughout the day, specified using the Emission Production Model developed by the U.S. Forest Service (http://www.fs.fed.us/pnw/fera/sue/epm). The emissions of species (as moles/s) are distributed into vertical layers according to plume rise formulas driven with MM5 meteorology. Daily total fire emissions in metric tons are shown in Figure 4-5. Major fires occurred in late July (e.g., the Happy fire), see Figure 4-4) and late August (e.g., the Bark fire, see Figure 4-4). During large fires, only one-third of the NOx emissions are allocated near the surface (within 200 m), while more than half of the VOC emissions are injected within the same height.

Figure 4-5: Daily Total Forest Fire Emissions for the CCOS Domain During Summer 2000



The Julian date range shown corresponds to 2 June to 27 September.

4.1.1 Meteorological Fields

The National Center for Atmospheric Research/Pennsylvania State University Mesoscale Model (MM5) (Grell et al. 1994) version 3 was used to perform a series of five-day simulations (except for two three-day periods at the beginning and end of the time period) starting at 1200 UTC 1 June to 30 September 2000 and centered around the five-day Intensive Observation Period (IOP) that began on 29 July 2000, which is the same as used in Michelson et al. (2009). The simulations for the whole summer period were carried out without data assimilation. A set of 36-12-4 km one-way nested grids that have 50 vertical stretched levels was used, 30 of which are within the lowest 2 km. The middle point of the lowest model layer is at about 12 meters above the surface. Boundary and initial conditions were prescribed using the six-hourly 40 km National Centers for Environmental Prediction (NCEP) Eta analyses. All of the simulations used the Eta Atmospheric boundary layer scheme (Janjic 2002). The Dudhia simple microphysics parameterization was used, along with the Dudhia short-wave and Rapid Radiation Transfer Model (RRTM) long-wave radiation parameterization schemes. The 36- and 12-km grids used the Grell convective parameterization scheme (Grell 1993), while no convective parameterization scheme was used on the 4-km grid. Details of all the physics options are available in Grell et al. (1994).

Meteorological fields with 4 km resolution were used as input for air quality modeling in this study. The Meteorology to Chemistry Interface Processor (MCIP version 3.3) was used to construct CMAQ model-ready input files from the MM5 output, and it allows for consolidation of vertical layers. We used 27 layers for CMAQ from the original 50 MM5 layers without changing the first 200 m (9 layers) to preserve high resolution near the ground. To avoid the potential influences of boundary grid cells, the origin of the meteorological domain was shifted inside by two grids, and a subset of the full 190 × 190 MM5 4-km domain was used for driving air quality simulations.

4.1.2 Intensive Operating Periods (IOPs)

The typical synoptic scale meteorological features favoring ozone formation and accumulation in Central California are strong high-pressure systems built over the eastern Pacific (Eastern Pacific High) or the Four Corners (Western U.S. High). The high-pressure system is manifest as a dome of warm air (a maximum in the 500-millibar geopotential height field) with a surrounding anticyclonic circulation (clockwise in the Northern Hemisphere). The sinking motion of air aloft is associated with fair weather and gives rise to adiabatic heating and warm temperatures. An offshore pressure gradient caused by the high-pressure system reduces the incoming westerly flow at the coast and creates stagnant conditions in the San Joaquin Valley. A less typical weather associated with high ozone levels is monsoonal flow: an upper-level high is centered in the southwestern United States or in northern Mexico such that flow from the south (a southerly flow) transports moisture north. The atmosphere is less stable in this case, but ozone exceedances in the SJV often result.

Several IOPs were identified during CCOS; each one is associated with one of the above synoptic patterns. High ozone concentrations were observed during all of these IOPs. Detailed

descriptions can be found in Fujita et al. (2001). We extend shorter IOPs (one or two days) to three days to facilitate the analysis.

1. Jul 23–25 (day 205–207): Monsoonal Flow
2. Jul 29–Aug 2 (day 211–215): Western U.S. High
3. Aug 14-17 (day 227-230): Western US High
4. Sept 17-20 (day 261-264): Eastern Pacific High

4.2 Air Quality Model Simulations and Performance Evaluation

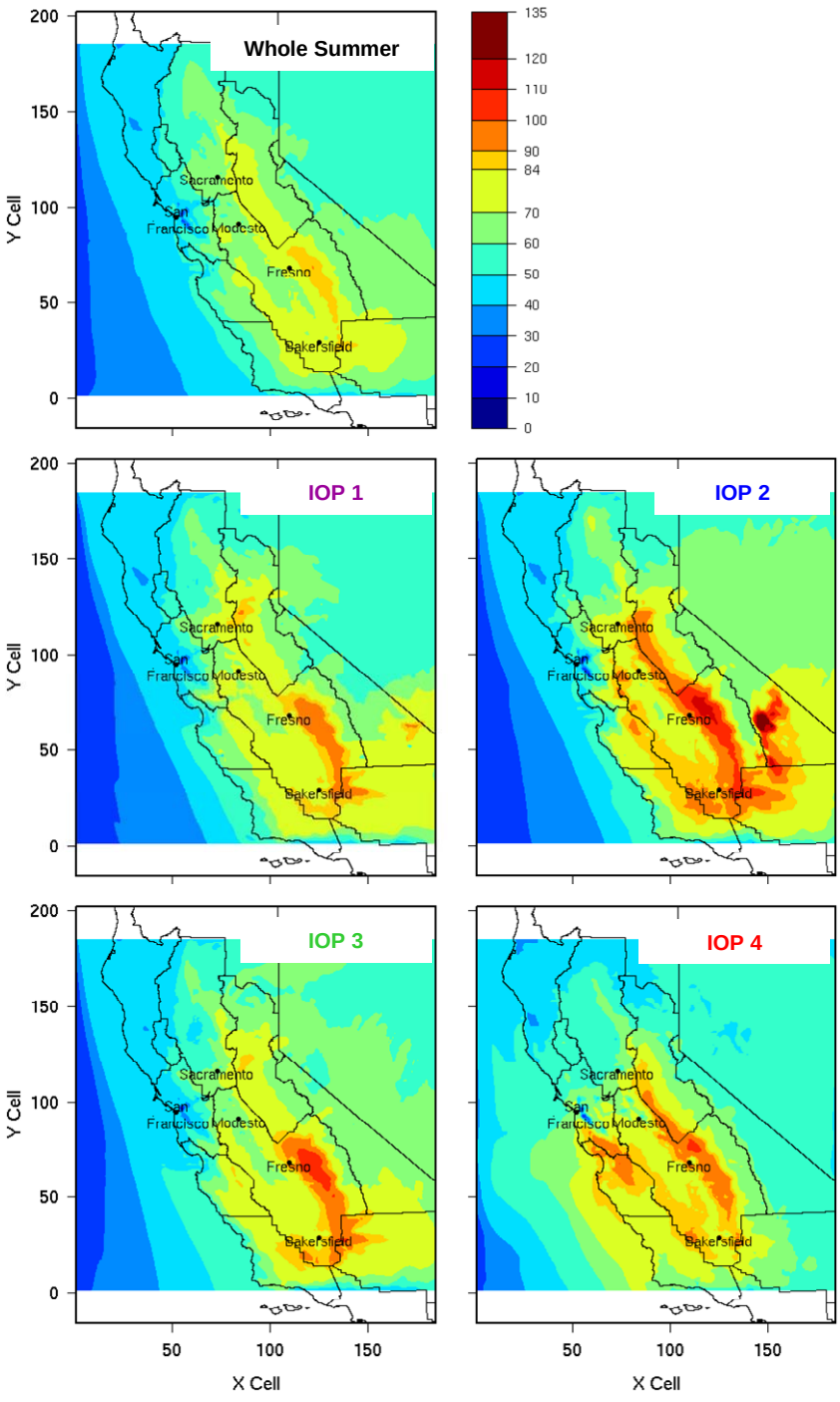
Version 4.5 of CMAQ was configured to use the piecewise parabolic method for advection, horizontal diffusion, and vertical diffusion, and the Euler Backward Iterative (EBI) ordinary differential equation solver for chemistry. Gas-phase chemistry is represented using the SAPRC99 chemical mechanism (Carter 2000). Other configurations, such as dry deposition velocity, vertical diffusivity, boundary conditions, and others, are the same as the revised parameters in Chapter 2.

4.2.1 Simulated Ozone Concentrations

Simulated 8-h ozone maxima averaged over the whole summer and for individual IOPs are shown in Figure 4-6, with ozone levels greater than 84 ppb ("high ozone") are marked in the darker shade of orange. During the summer, on average, high ozone levels occur downwind of Fresno and Bakersfield. Four IOPs each showed distinct ozone spatial patterns. Coastal areas and the SFB in IOP1 were relatively clean, while high ozone was found in the Mountain counties and eastern side of the SJV. IOP2 exhibited the highest domain-wide ozone levels among all the IOPs, with elevated ozone in Sacramento, both sides of the SJV, the Mountain counties, and near the coastal mountain ranges. Elevated ozone levels in IOP3 were mostly seen in the SJV; in IOP4, high ozone levels were seen in the SFB and downwind coastal areas, as well as in the SJV and the Mountain counties. In all cases, ozone levels in the SJV were high.

Hourly ozone concentrations were measured at 149 ground stations located throughout the CCOS domain. At individual sites, modeled ozone time series followed the observed temporal patterns (time series plots can be provided upon request). The model is able to capture both high- and lower-ozone days. Nighttime ozone levels were often over-predicted, a problem that has been encountered in other modeling studies (e.g., Eder 2006). The nighttime discrepancies are partly due to the titration of O_3 with fresh NOx emissions. Additionally, nighttime ozone levels are more prone to the uncertainties in vertical mixing as we discussed in our preliminary diagnostic simulations (Chapter 2). Hence, our evaluation of ozone focuses on daily 1-h and 8-h maxima, and when comparing hourly ozone levels, a cutoff value of 60 ppb will be used so as to focus on daytime ozone when vertical mixing is more vigorous. Since nighttime concentrations will be affected by titration effects, model-predicted total nitrogen species (NOy) will only be compared to observations during the better-mixed daytime hours (10 a.m.~5 p.m., when the PBL height remains high).

Figure 4-6: Simulated 8-h Peak Ozone (ppb)



4.2.2 Temporal and Spatial Behaviors of Ozone Predictions

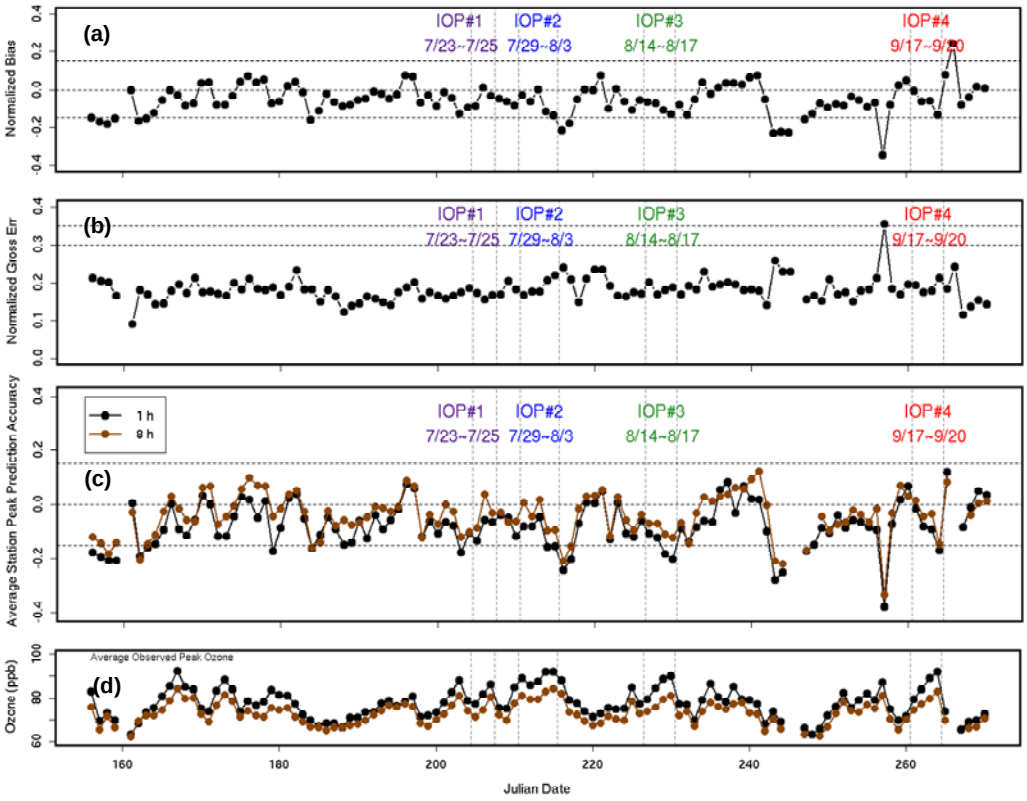
To investigate ozone model performance further, daily normalized biases and gross errors were calculated (Figure 4-7ab) using the hours and sites where observed ozone levels are greater than

60 ppb. The normalized biases are largely within ±15 percent throughout the whole summer period, with only a few exceptions. The normalized gross errors are mostly less than 25 percent, and only one day is more than 35 percent.

Daily average station peak prediction accuracy[7] is calculated for 1-h and 8-h average peak ozone (Figure 4-7c). Observed peak ozone is determined as the maximum value in the time interval from 10 a.m. to 8 p.m. without more than three missing values. Peak prediction accuracy was measured with strict pairing in space and time.

The model generally underpredicts peak ozone (for observed peak > 60 ppb), especially for the 1-h peak. Figure 4-7d shows observed daily average peak ozone. One can see that the model underpredicts peak ozone on both high and low ozone days.

Figure 4-7: Ozone and Evaluation Metrics by Day



(a) Daily normalized biases and (b) normalized gross errors of hourly ozone with cutoff = 60 ppb; (c) daily average station peak prediction accuracy (for 1-h and 8-h peak ozone), along with (d) daily average observed peak ozone levels. Horizontal dashed lines denote U.S. EPA guidance for acceptable model performance (normalized bias within ±15%, and normalized gross error less than 30%–35%).

[7] Defined by U.S. EPA model evaluation guidance (1991), equivalent to peak ozone biases averaged over all the measurement stations in the domain.

To evaluate the spatial behavior in model predictions, ozone performance statistics are aggregated over the whole domain and for selected air basins where high ozone levels are observed, namely in the SJV, SV, SFB, and MC (Table 4-1). Despite the spatial differences, across all the air basins, the summer average metrics for ozone (hourly or peak) only showed modest underprediction within 10 percent and normalized gross errors around 20 percent. These statistics are better than the ones over California found by previous studies (Tong and Mauzerall 2006; Eder et al. 2009) conducted with coarser horizontal resolutions. Ozone predictions in the SJV compare better with observations than for the CCOS domain as a whole, while the SFB generally has less satisfactory performance than other subregions. Initial model diagnostics showed that the SFB is highly sensitive to uncertainties in both anthropogenic and biogenic emission sources, boundary conditions, sea breeze strength and timing, and light attenuation by coastal stratus, which together pose significant challenges for simulating ozone concentrations.

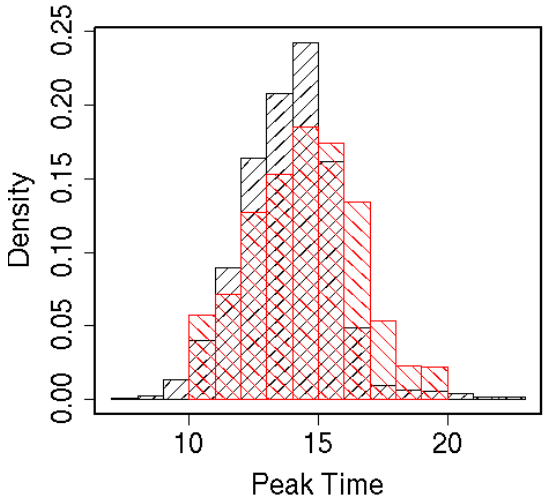
Table 4-1: Performance Statistics.

Cutoff =60 ppb	Hourly Ozone					1-h Peak Ozone					8-h Peak Ozone				
	Bias	N.Bias	Gross Err	N. Gross Err	RMSE	Bias	N.Bias	Gross Err	N. Gross Err	RMSE	Bias	N.Bias	Gross Err	N. Gross Err	RMSE
Domain	-4	-5%	14	19%	18	-7	-8%	16	19%	21	-4	-4%	12	15%	15
SFB	-7	-10%	18	26%	23	-4	-4%	17	22%	22	-7	-9%	13	18%	17
SV	-3	-4%	15	21%	19	-4	-4%	15	20%	20	-2	-2%	11	16%	14
SJV	-2	-2%	13	16%	17	-6	-5%	14	16%	18	-1	0%	10	13%	13
MC	-3	-4%	13	17%	16	-5	-6%	15	18%	19	-2	-1%	11	15%	14

Shown for the entire Central California study domain, and for selected subregions (SJV, SV, SFB, and MC).

Averaged over the entire summer, the model underpredicts ozone—including hourly values, as well as 1-h and 8-h peak ozone, in all the subregions we examined—for cases with observed ozone levels greater than 60 ppb. Strict pairing in time and space can be unrealistic given the uncertainties in the input data and model processes. Part of the problem with 1-h peak values may be a mismatch in timing of modeled versus observed peak ozone. When the comparison of simulated peak values is relaxed to be within a 3-h window centered on the observed peak hour, the average bias of 1-h peak ozone changed from negative to positive across all the air basins and (normalized) gross errors were reduced or remained unchanged, except for the SJV, where the gross error increased slightly but is still within 25 percent. According to the distributions of hourly ozone peak time (Figure 4-8), only 8 percent of the modeled peaks occur after 4 p.m., compared to 23 percent of the observed peaks. Ozone tends to decrease during the transition hours (5–7 p.m.) before sunset due to reduced photolysis rates. Further inspection of diurnal ozone profiles indicates that this decreasing trend occurs earlier and/or is steeper in the model than in the observations. This may suggest that the modeled photolysis rates are too low in these hours. The meteorological evaluation in the next section suggests that the transition hours are subject to more uncertainties in predicted boundary layer heights and wind fields, which may also affect simulated ozone concentrations. Predicted 8-h average peak ozone compares better with observed values than do the 1-h peak concentrations.

Figure 4-8: Distribution of 1-h Ozone Peak Time



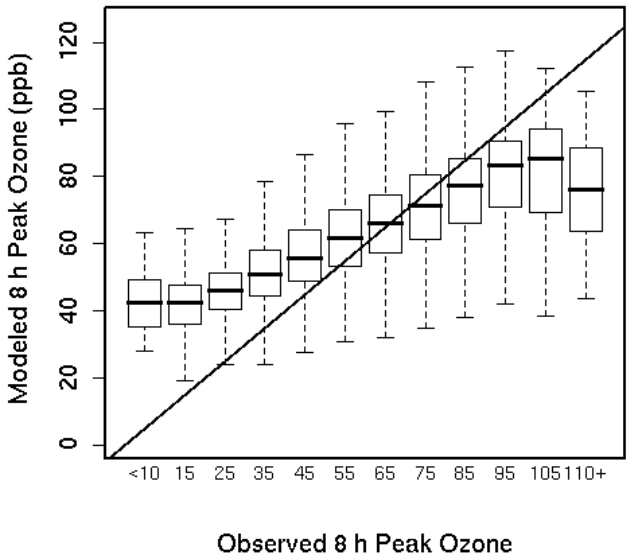
Cutoff = 60 ppb. Red: observed, black: modeled. Modeled peak is constrained to be within ±3 h of the observed peak.

Relaxing the spatial pairing of modeled ozone to be within ±1, 3, or 5 grid cell(s) of the observation site generally does not improve model performance statistics for either 1-h or 8-h

peak values. Subgrid differences in pairing are not investigated here; however, it has been found that the incommensurability between model volume-averages and the pointwise observations may contribute to poorer performance statistics when ozone spatial correlation assumes sharper decreases for short distances (Swall and Foley 2009).

Performance statistics are sensitive to the choice of cutoff concentration, the value below which observations are excluded in evaluating model performance. When the cutoff concentration for ozone is reduced from 60 to 40 ppb, the biases become positive: ~2 ppb for hourly ozone, and 2–6 ppb for 8-h peak ozone, across all subregions. Figure 4-9 shows the comparison with the complete range of observed 8-h peak ozone. The model tends to overpredict ozone in the lower range (< 60 ppb), which is similar to findings from other studies (Tong and Mauzerall 2006; Eder et al. 2006). The model predicts more accurately in the 60 to 80 ppb range, while other studies found the best prediction range is in the lower concentrations (40~50 ppb in Tong and Mauzerall 2006, and 50~70 ppb in Eder et al. 2009). Eder et al. (2006) attributed overpredictions of the smaller values of observed 8-h peak ozone to days with non-conducive meteorological conditions that have more cloud cover and/or precipitation. In Central California, summer time is characterized by clear skies and low humidity. Only in the coastal regions, where ozone levels are consistently low, do we observe ozone overpredictions throughout the summer. Spatial distributions of normalized biases at individual grid cells are examined in the next section.

Figure 4-9: Comparison of Modeled and Observed 8-h Peak Ozone



Distribution of modeled ozone corresponding to observed ozone levels (binned by 10 ppb) is shown with box plots farthest point within 1.5 times of the inter-quartile length). (showing median, 25th, and 75th percentile, and whiskers extend to the

4.2.3 Model Performance for Different Synoptic Events

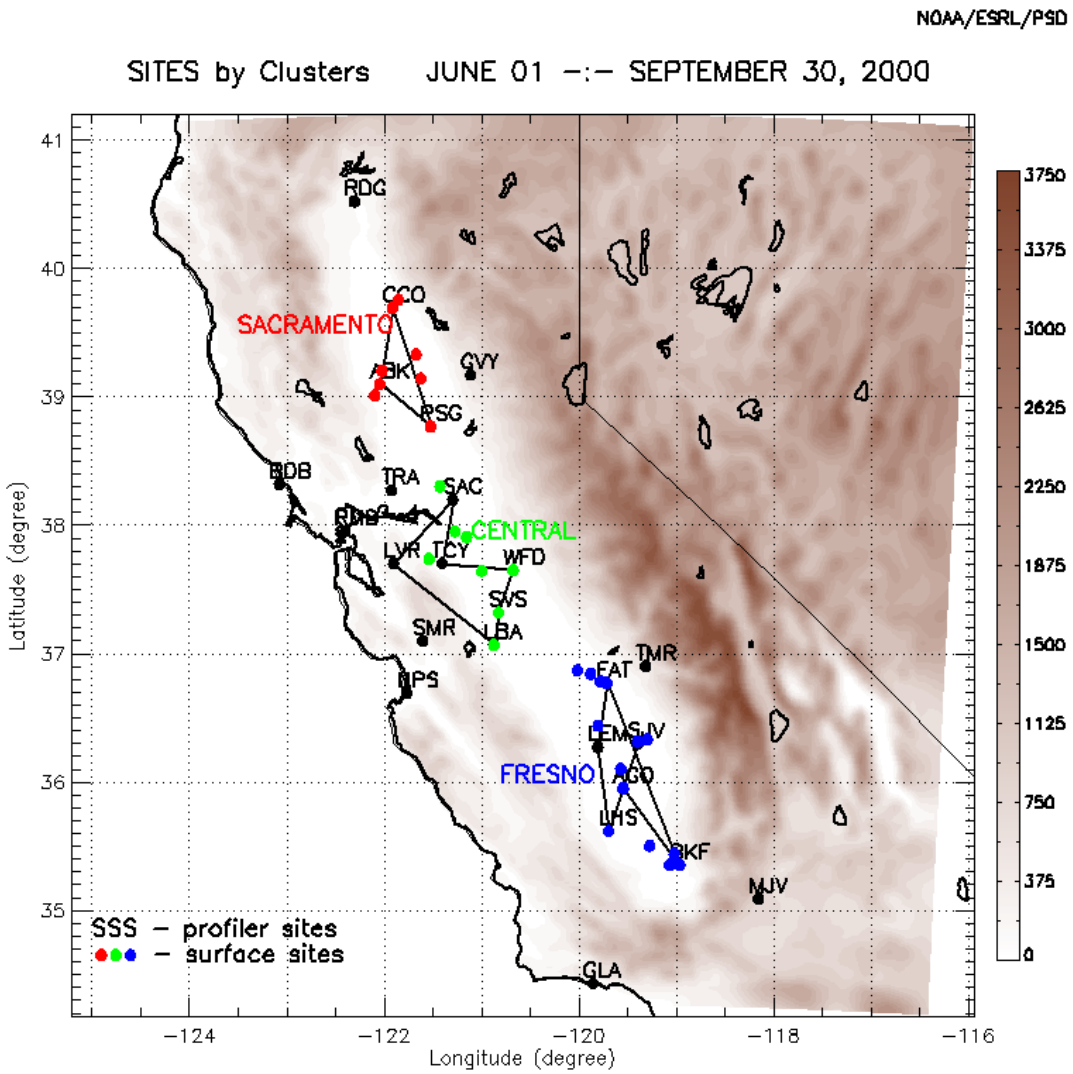
To identify common and distinctive features in model performance under different meteorological conditions, we evaluate both meteorological and chemical simulations across synoptic events (exemplified by the four IOPs) and compare them with average model performance for the whole summer.

4.2.3.1 Evaluation of Meteorological Parameters

The meteorological fields are evaluated by dividing the Central Valley (consisting of Sacramento Valley and San Joaquin Valley) into three areas: the (middle) Sacramento Valley (SV), the central part, and the (middle-southern) San Joaquin Valley (SJV). Three areal clusters of profiler locations are used to evaluate the model performance in each of these regions. These clusters are referred to as the *Sacramento cluster,* the *Central cluster,* and the *Fresno cluster,* respectively (see Figure 4-10).

The wind-profiler sites that are included in the averaging in the Sacramento cluster are: Chico (CCO), Arbuckle (ABK), and Pleasant Grove (PSG). The profiler sites included in the Central cluster are: Travis (TRA), Livermore (LVR), Los Banos (LBA), Waterford (WFD), Tracy (TCY), and Sacramento (SAC). The profiler sites included in the Fresno cluster are: Fresno (FAT), Lemoore (LEM), Lost Hills (LHS), Angiola (AGO), Bakersfield (BKF), and Visalia (shown as SJV in Figure 4-10 below). The wind profilers provide hourly averages of wind speed and direction, typically to heights of 3,000 m above ground level (AGL), and virtual temperature up to about 1,000 m AGL determined using the Radio Acoustic Sounding System technique. The information collected by the wind profilers can also be used to determine the depth of the daytime convective atmospheric boundary layer (ABL) by visually inspecting values of range-corrected signal-to-noise ratio, vertical velocity (which is large within the convective boundary layer), and radar spectral width (which is a measure of turbulence intensity). Meteorological parameters examined here are wind speed and direction, ABL depth, and surface temperature.

Figure 4-10: Location of Wind Profiler Clusters Used for MM5 Model Evaluation

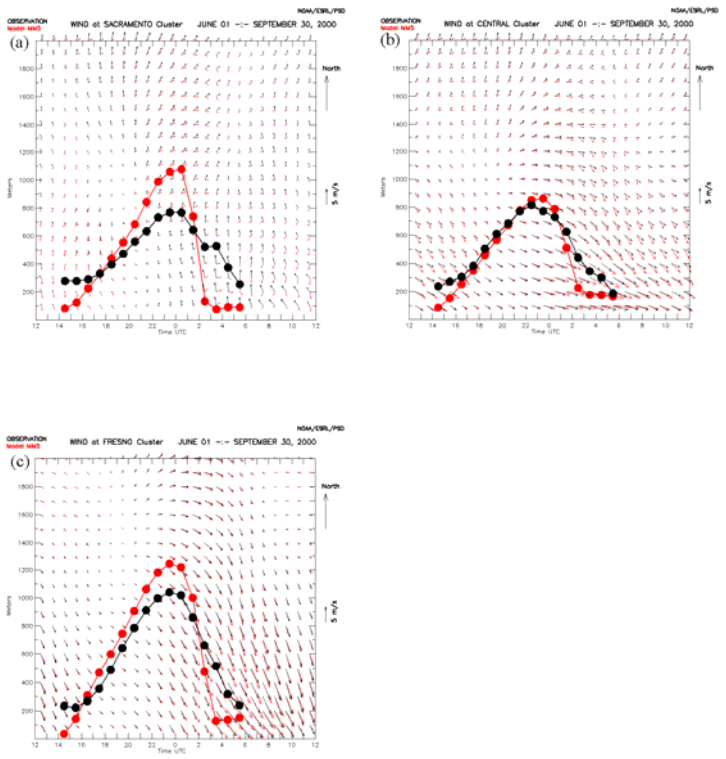


The Sacramento cluster is shown in red, the Central cluster in green, and the Fresno cluster in blue.

Figure 4-11 shows a time height series of the diurnal cycle of MM5-simulated and profiler-derived winds averaged over the entire summertime period in the three areal clusters (Sacramento, Central, and Fresno clusters). The Sacramento cluster characterizes meteorology in the Sacramento Valley. The central cluster characterizes the incoming marine flows from the Bay area to the Central Valley, which directly influence the transport and accumulation of

pollutants in the Bay area and its intermediate downwind regions (north central coasts, Sacramento metropolitan area, and northern San Joaquin Valley). The Fresno cluster covers the middle to southern parts of the San Joaquin Valley. The MM5-simulated winds in the Fresno cluster are in the best agreement with the profiler-observed winds at all levels and times. Errors in the wind direction are greatest in the Sacramento cluster. The winds within the daytime ABL are simulated well, better than any other time and any other height, in both the Central and Fresno clusters. This good agreement of the winds in the ABL indicates that the horizontal transport within the daytime ABL is simulated well in the SJV. The biggest differences in the Fresno cluster occur at night, when the low-level jet that is present between 0000 and 0800 UTC (5 p.m.~1 a.m. Pacific Daylight Time) and below about 1,000 m, is too strong in the MM5 simulations. The simulated depth of the ABL is larger than observed in all three clusters during the well-mixed hours (10 a.m.–5 p.m.). During the transition hours in the afternoon (5 p.m.–8 p.m.), the lower MM5 ABL height indicates less turbulent mixing in the model than that observed in all three clusters. Further discussion of the reasons for differences between the observed and simulated ABL is available in Michelson et al. (2009).

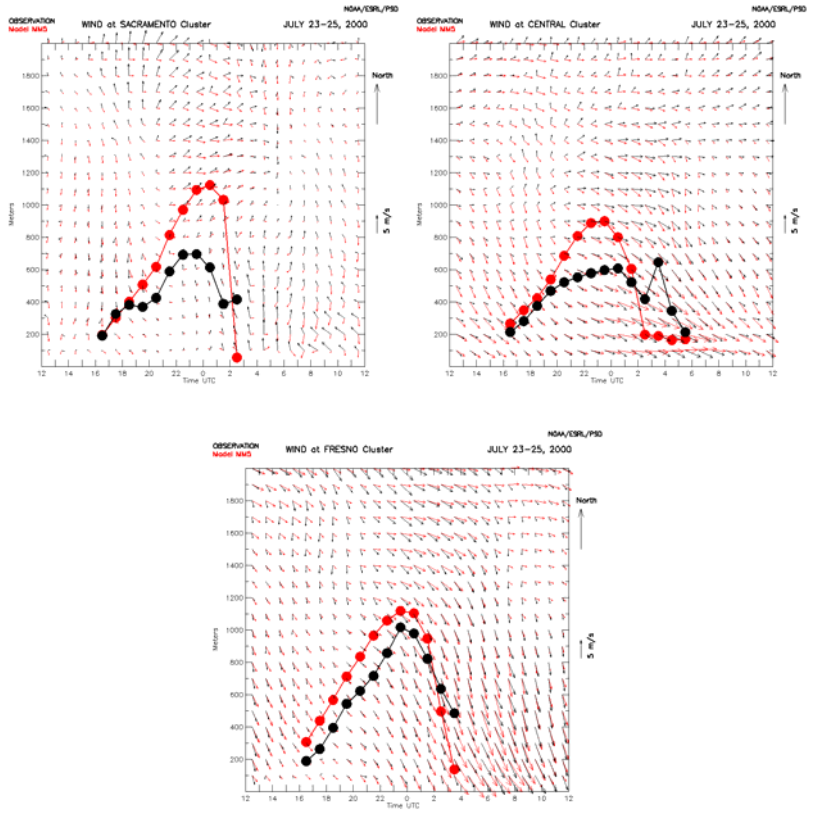
Figure 4-11: Time-Height Series Plot for Meteorological Evaluation: Summer



MM5-simulated (red vectors) and observed (black vectors) wind along with the MM5-simulated (red dots) and observed (black dots) atmospheric boundary layer depth averaged over entire summer (1 June–30 September 2000).

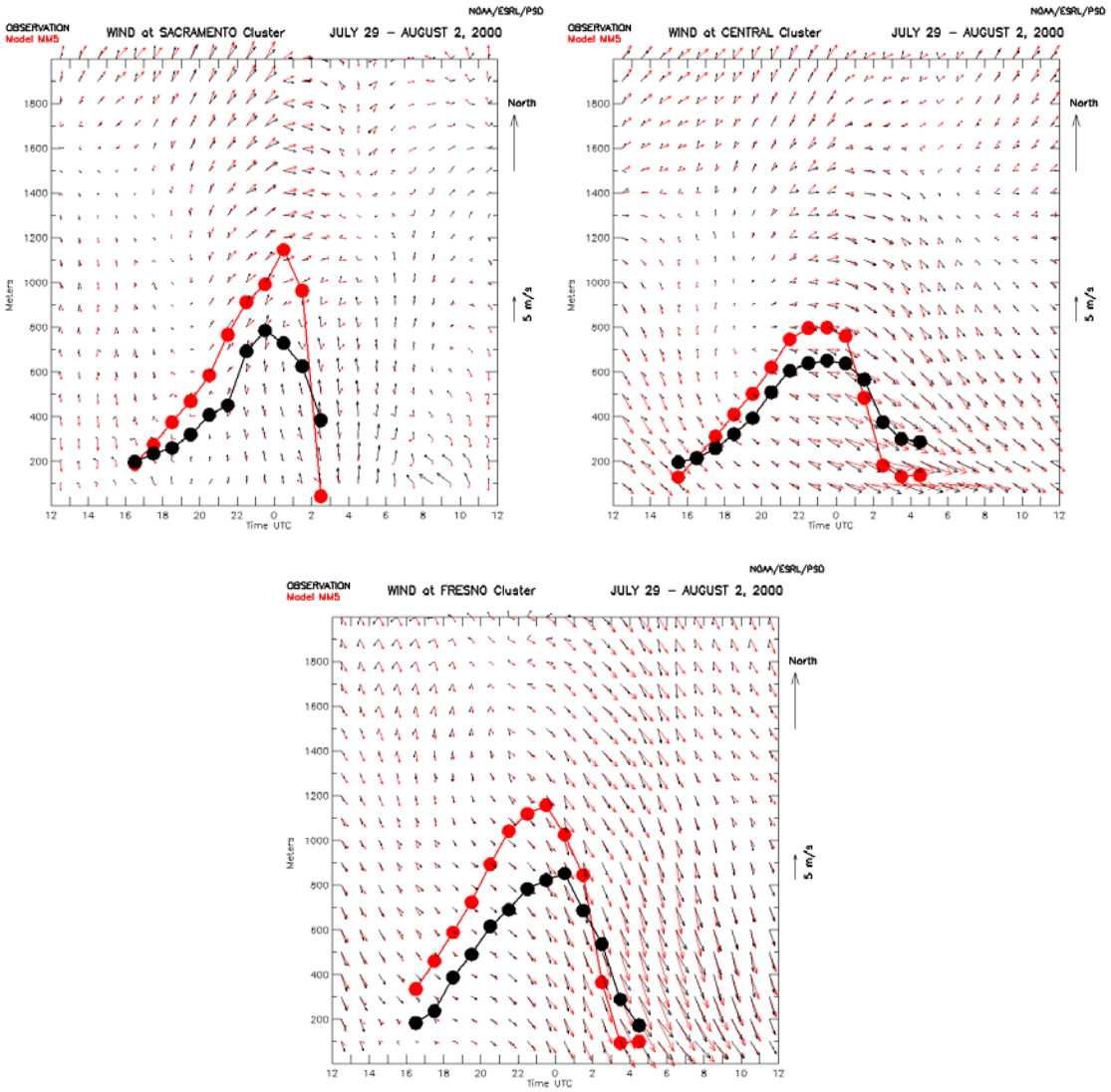
Similar assessment of simulated winds and ABL depth is conducted for the four IOPs (see figures 4-12 to 4-15). In common with the summer average, the simulated depth of daytime ABL in individual IOPs has a positive bias, except for the last IOP in September. The averaged diurnal cycle of winds in IOP4 is different from the other three IOPs and the summer average, because the incoming marine flow in the Central cluster and the nocturnal low-level jets in the Fresno cluster in IOP4 are much weaker than the other cases. The Sacramento cluster exhibits the poorest agreement in wind directions in ABL during IOP3, and the simulated northwesterly component and the overall wind speeds are too strong during IOP4. Wind features averaged over the Central cluster characterize the incoming marine flows (indicated by the westerly component). During IOP3, the simulated marine flow is stronger than observed. During the more stagnant conditions in IOP4, the simulated marine flow is also stronger in the morning and directed more toward the south. The Fresno cluster generally has good agreement between simulated and observed winds within the daytime ABL, and performance in each IOP is similar to the summer average. Temperature evaluations can be found in Michelson et al. (2009). In general, MM5 underpredicts surface temperature by about 2°C to 3°C (4°F to 5°F).

Figure 4-12: Time-Height Series Plot for Meteorological Evaluation: IOP 1



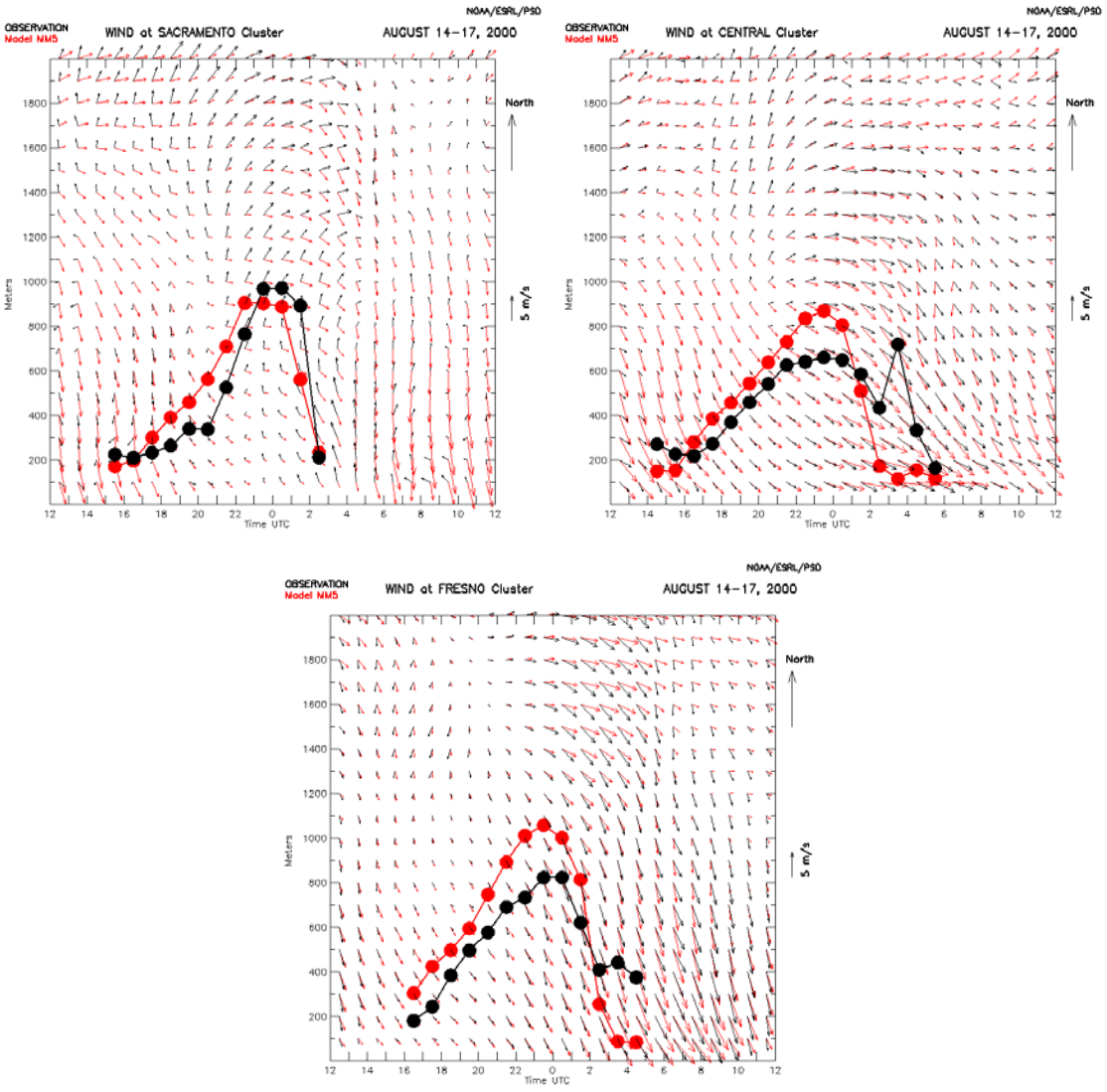
MM5-simulated (red vectors) and observed (black vectors) wind along with the MM5-simulated (red dots) and observed (black dots) atmospheric boundary layer depths averaged over IOP 1, 23–25 July 2000.

Figure 4-13: Time-Height Series Plot for Meteorological Evaluation: IOP 2



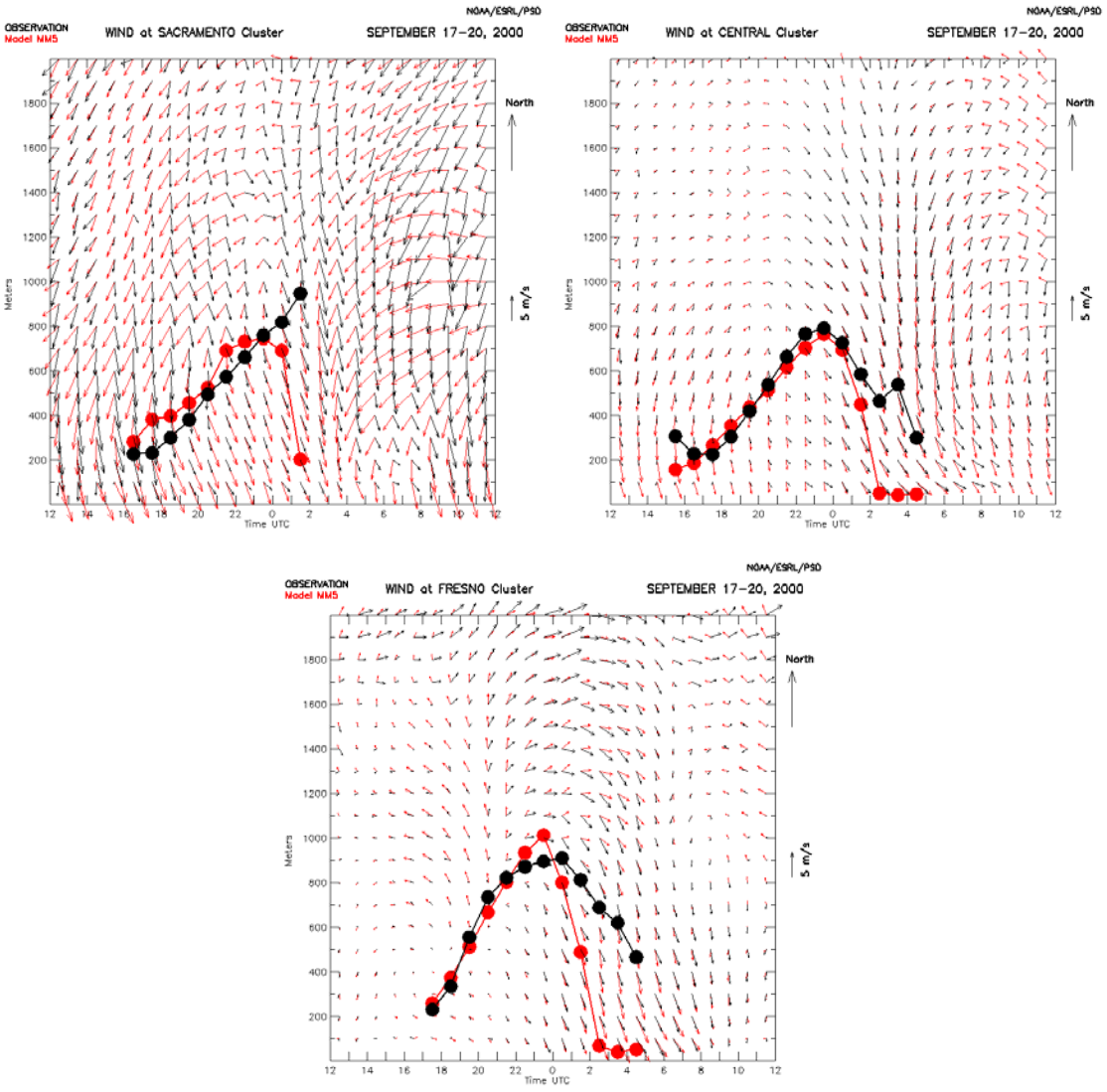
MM5-simulated (red vectors) and observed (black vectors) wind along with the MM5-simulated (red dots) and observed (black dots) atmospheric boundary layer depths averaged over IOP 2, 29 July–3 August 2000.

Figure 4-14: Time-Height Series Plot for Meteorological Evaluation: IOP 3



MM5-simulated (red vectors) and observed (black vectors) wind along with the MM5-simulated (red dots) and observed (black dots) atmospheric boundary layer depths averaged over IOP 3, 14–17 August 2000.

Figure 4-15: Time-Height Series Plot for Meteorological Evaluation: IOP 4



MM5-simulated (red vectors) and observed (black vectors) wind along with the MM5-simulated (red dots) and observed (black dots) atmospheric boundary layer depths averaged over IOP 4, 17–20 September 2000.

Wind speed and direction within the mixing layer are especially important for determining ozone transport and accumulation in particular locations. The mixing height determined from MM5 is used to calculate vertical diffusivity in CMAQ. Concentrations of primary pollutants are directly affected by both wind speed and mixing height, through advection and turbulent diffusion, and greater dilution is associated with lower concentrations. For secondary pollutants like ozone, in addition to reducing concentrations by direct transport, dilution rates can affect production rates differently and this depends on localized chemical regimes. In NOx-rich locations, reduced NOx concentrations will decrease titration and increase ozone concentrations. This effect may offset the direct dilution effect and lead to a net increase in

ozone concentrations. We perturbed daytime (10 a.m. to 5 p.m.) ABL depth by 20 percent and verified that changes in ozone can be positive or negative, depending on the location. Dilution rates also affect ozone sensitivities. As elaborated in Biswas et al. (2001), more rapid dilution of pollutants tends to shift local chemistry in polluted plumes (usually NOx-rich) toward the NOx limited regime.

Uncertainties in temperature have more complicated effects on ozone chemistry,[8] as the rate coefficients may increase or decrease depending on the reaction (Steiner et al. 2006). Most reaction rates increase with higher temperature; however, some three-body reactions proceed more slowly. A simulation with a 2ºC temperature increase was conducted and it showed ozone increases in the domain with maximum changes in the regions with high anthropogenic emissions such as the Bay area, Fresno, and Bakersfield, which is similar to the findings of Steiner et al. (2006). Ozone sensitivity to anthropogenic emissions was also evaluated with and without temperature increases, using the Decoupled Direct Method (DDM). The most noticeable differences in ozone sensitivities were found in the Bay area, where the largest ozone changes with temperature occur. The Bay area is in a VOC-limited regime, and ozone sensitivity to NOx decreases with temperature, while ozone sensitivity to VOC increases, because more radicals are generated. In summary, a negative bias in temperature generally leads to a decrease in ozone concentrations, decreased sensitivity to VOC, and increased sensitivity to NOx.

4.2.3.2 Evaluation of CMAQ Simulations

Evaluation metrics regarding chemical model outputs are also produced for the summer season and individual IOPs.

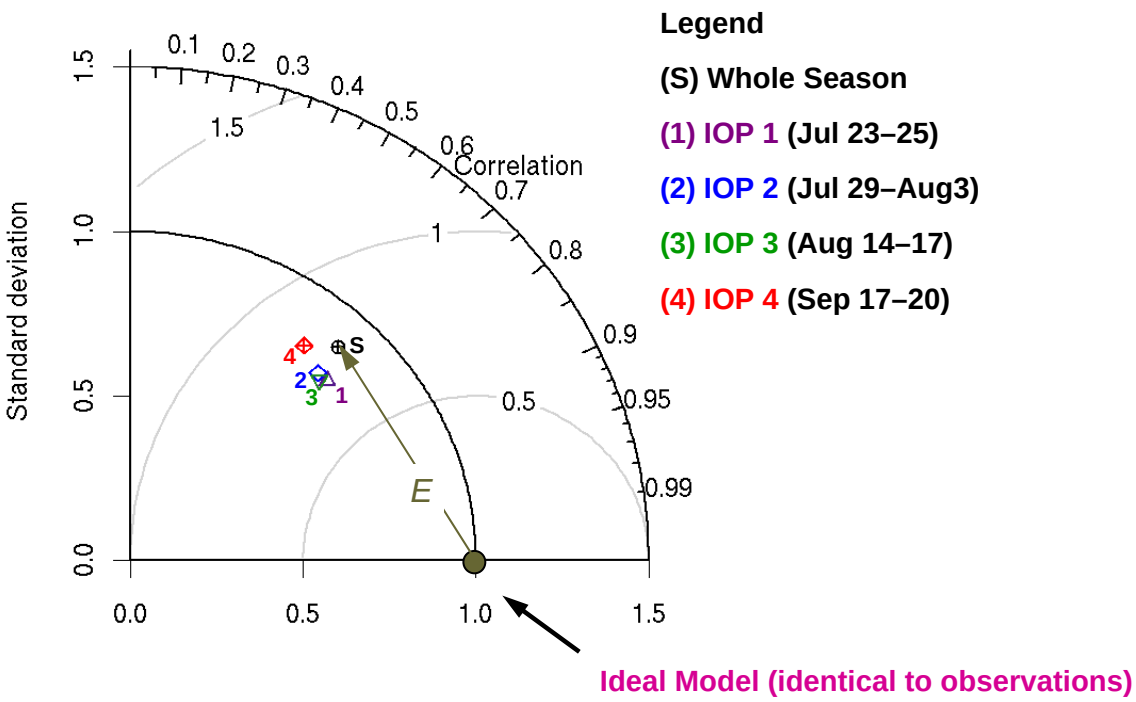
4.2.3.2.1 Ozone Predictions

A Taylor diagram, as described in Chapter 2, is employed here to evaluate modeled ozone patterns. This method has been applied to evaluation of climate and air quality models (e.g., Vautard et al. 2007).

We examine daytime (6 a.m.–9 p.m.) ozone patterns captured by the model in the four IOPs compared with the entire summer season (Figure 4-16). The whole summer season and four IOPs form a tight cluster on the Taylor diagram. This indicates that there are no large differences in model performance between the IOPs and the whole summer in terms of variability measured by the three pattern statistics. Modeled daytime ozone in all the IOPs, as well as the whole summer on average, tend to be less variable (as measured by standard deviation) than the observations. Correlations between the modeled and observed ozone levels are in the range of 0.6–0.7 across all time periods. All IOPs have similar root-mean-square differences (~0.75 of the observed standard deviation), and model performance for IOP4 is slightly inferior to the other IOPs (i.e., lower correlation and larger RMS error).

[8] Uncertainties in simulated temperatures here are uncoupled from emissions, which are estimated using observed temperatures.

Figure 4-16: Pattern Statistics for IOPs and the Whole Season



The Taylor diagram indicates three pattern statistics: standard deviation, root-mean-square difference (E indicated on the diagram), and correlation between the modeled and observed daytime ozone for the whole summer season and individual intensive operating periods. Standard deviation and root-mean-square difference are normalized using the standard deviation of the observations.

Pattern statistics are useful for comparing the anomalies in two data sets, while the mean bias and gross error employed earlier indicate the sign and magnitude of model prediction errors. Modeled hourly ozone, and peak ozone (1-h and 8-h) greater than 60 ppb, are under-estimated by the model across all the time periods, as shown in Table 4-2. IOP 3 has the largest ozone underpredictions of all the IOPs indicated by both bias metrics, with a normalized bias that is twice the summer average. Other statistical measures do not suggest that model performance for IOP 3 is inferior to the others.

Table 4-2: Summary of Model Performance

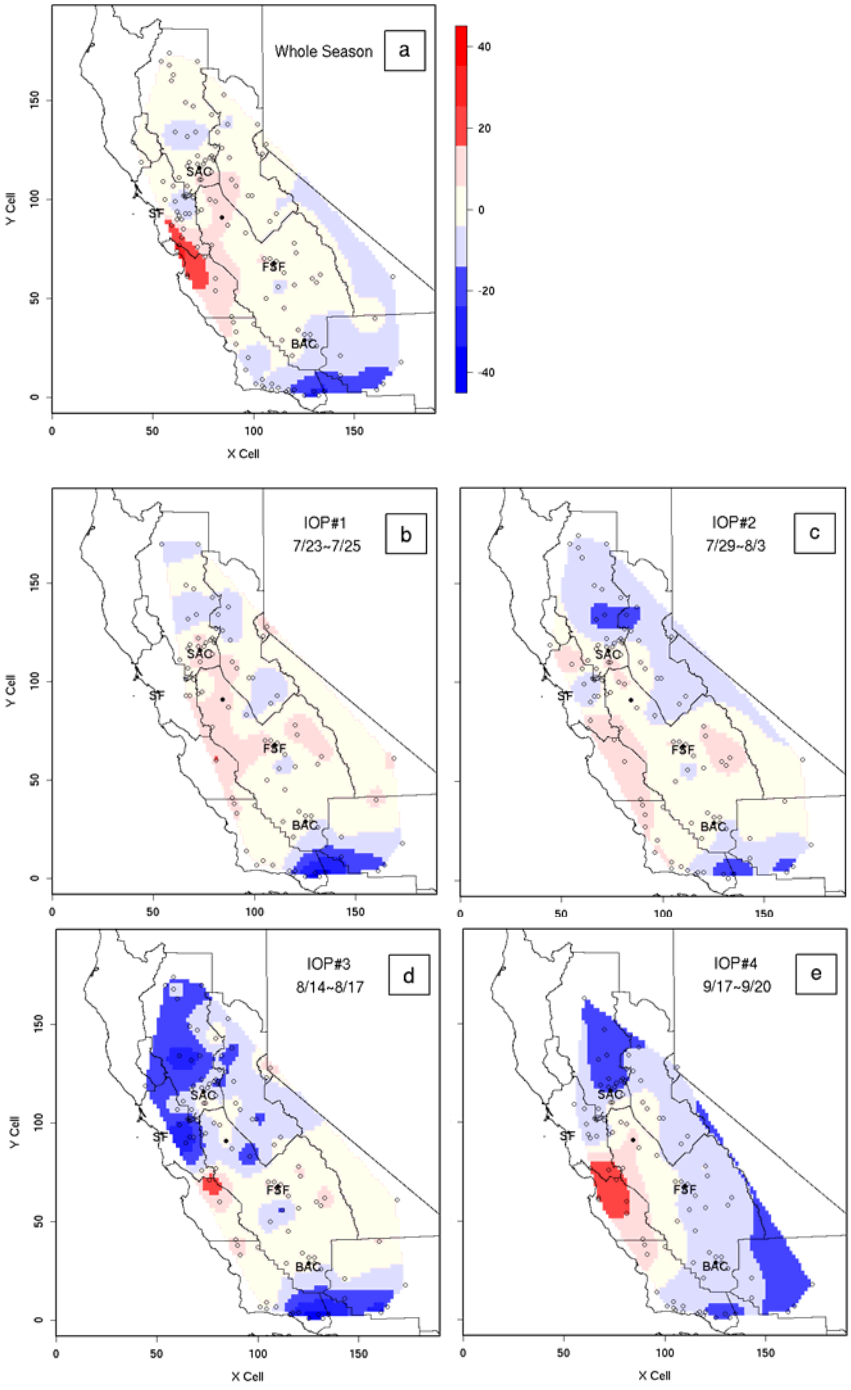
	Hourly Ozone (cutoff = 60 ppb)					1-h Peak Ozone (cutoff = 60 ppb)				
	Summer	IOP 1	IOP 2	IOP 3	IOP 4	Summer	IOP 1	IOP 2	IOP 3	IOP 4
Bias	-4	-3	-6	-8	-6	-7	-8	-11	-14	-10
N. Bias	-5%	-4%	-7%	-10%	-7%	-8%	-9%	-11%	-16%	-9%
Gross Err	14	12	15	14	15	16	15	18	18	18
N. Gross Err	19%	17%	19%	18%	19%	19%	18%	20%	20%	20%
RMSE	18	17	20	19	20	21	19	23	24	23
Sample Size	59,158	2,245	5,046	3,588	2,638	7,578	266	511	406	364
	8-h Peak Ozone (cutoff = 60 ppb)					NOy (cutoff = 5 ppb)				
	Summer	IOP 1	IOP 2	IOP 3	IOP 4	Summer	IOP 1	IOP 2	IOP 3	IOP 4
Bias	-4	-3	-5	-8	-6	-1	-4	-6	-3	-2
N. Bias	-4%	-3%	-5%	-9%	-6%	3%	-9%	-18%	-17%	-5%
Gross Err	12	10	12	12	13	6	7	8	6	6
N. Gross Err	15%	13%	15%	16%	17%	49%	43%	43%	41%	47%
RMSE	15	13	16	16	17	11	16	19	9	10
Sample Size	5,464	214	440	337	268	8,458	275	517	417	415
	NMHC (cutoff = 35 ppbC)					CO (cutoff = 200 ppb)				
	Summer	IOP 1	IOP 2	IOP 3	IOP 4	Summer	IOP 1	IOP 2	IOP 3	IOP 4
Bias	-4	5	-14	-9	2	-184	-119	-238	-540	-289
N. Bias	23%	20%	23%	-7%	36%	-31%	-26%	-29%	-52%	-31%
Gross Err	66	53	88	22	68	254	187	320	573	367
N. Gross Err	63%	69%	70%	33%	66%	46%	41%	47%	62%	48%
RMSE	128	103	177	29	110	399	279	465	660	602
Sample Size	195	39	59	16	69	11,4407	3,746	5,359	2,601	5,139

In summary, the domain-wide aggregated metrics (both pattern statistics and other metrics) generally indicate similar performance between the IOPs and the whole summer with respect to predicted ozone. Next we will see that episode-to-episode differences in ozone predictions are more pronounced at a subregional level.

Figure 4-17 shows the spatial distribution of normalized biases in 8-h peak ozone (cutoff = 60 ppb) averaged over the whole summer and for individual IOPs. Large negative biases seen at the southern boundary sites across all the periods are due to the influence of ozone lateral boundary conditions specified at the surface (40 ppb). Flux calculations indicated that the southern boundary is dominated by outflow (Harley et al. 2006), and sensitivity studies revealed that the lateral chemical boundary conditions here have little effect on the inner parts of the domain. Model performance is most robust (normalized biases within ±15 percent) in the SJV across all the time periods, even during IOP1 with its less typical monsoonal flows. Larger discrepancies between modeled and observed 8-h ozone maxima are seen in IOPs 3 and 4 in the Sacramento Valley, Bay area, and the North Central Coast, where the normalized biases deviate from the summer averages. The MM5 performance evaluation indicated that the winds in the ABL exhibited less agreement during IOP3 and IOP4 in the Sacramento and central clusters (especially from the morning to the early afternoon), and this is reflected in the ozone predictions. More specifically, in IOP3, simulated winds in the Sacramento Valley are too strong in the morning, and incorrectly show southerly flow in the afternoon, preventing transport of pollutants from Sacramento to the north. As a result, ozone is underpredicted in the northern part of the Sacramento Valley. At the same time, stronger northwesterly components in the simulated winds enhance the removal of pollutants from the Bay area and lead to an overprediction of ozone in downwind areas (the red coastal areas in Figure 4-17). In IOP4, simulated wind speeds are too high, and this dilutes pollutant concentrations in the Sacramento Valley. The simulated marine flow is directed more to the south, which reduces transport of SFB emissions into the Central Valley, causing significant accumulation of ozone along the coast.

Eder et al. (2006 and 2009) found that poorer model performance generally occurred under conditions that were not conducive to ozone (i.e., cloud cover, precipitation, and cool temperatures). In this study, only IOP1 occurred under a less typical Monsoonal Flow event, which has comparable model performance with summer averages. IOPs 2–4 in this study are driven by high-pressure systems, which are characterized by high temperature and low humidity. We have seen the poorest model performance in the SFB and/or SV during IOPs 3 and 4, which are subject to the largest uncertainties in the meteorological fields. This indicates that in Central California, the model performance is specific to subregions, where accurate simulations of meteorological fields are challenging even under the conducive conditions. Better seasonally averaged model performance in the SFB and SV may suggest temporal compensating errors in these regions.

Figure 4-17: Normalized Bias Surface of 8-h Peak Ozone Predictions



Shown for the whole summer (a) and individual IOPs (b–e). Cutoff value is 60 ppb. Biases are kriged into a continuous surface with the Matern function (Stein 1999) in R. Dots indicate surface observation sites. Major cities are labeled on the plots: SF (San Francisco), SAC (Sacramento), FSF (Fresno), and BAC (Bakersfield).

4.2.3.2.2 Precursor Predictions

In addition to ozone, the model's ability to simulate a variety of ozone precursors, including NOy, CO, and NMHC is evaluated. Due to the possible titration of ozone near large NOx sources under less well-mixed conditions, for NOy comparison, we only consider data in the interval from 10 a.m. to 5 p.m., when the atmospheric mixing is strongest, as indicated by the boundary layer heights. Modeled NMHC species are summed after being multiplied by their respective carbon numbers, and compared with observed NMHC in ppbC units.

When evaluating predictions of ozone precursors, a cutoff level corresponding to the 25th percentile of the observed values was used; results are presented in Table 4-2. On average, the model underpredicts NOy concentrations during daytime hours (mean biases -2 to -6 ppb). Normalized gross errors (40~50 percent) suggest similar model performance throughout the summer, while other metrics suggest that model performance for IOP2 is slightly less satisfactory than others. Predicted NMHC concentrations are also similar, except that IOP3 has only half of the normalized error of the other periods (60~70 percent), as well as a smaller normalized bias. This time period also has less than half of the number of observations compared to the other IOPs, and thus is not as representative. Carbon monoxide concentrations are underestimated (by ~30 percent, except IOP 3) across all time periods considered. Predictions during IOP3 compare poorly with the observations. This IOP occurred during August, when CO observations in the CCOS database tended to be rounded to the nearest 1 ppm, introducing an undesirable measurement roundoff error. In summary, these domain-wide metrics do not suggest that any episode is particularly worse than others, and precursor predictions generally have similar errors and biases among the IOPs and for the whole summer.

Normalized gross errors suggest that CMAQ does a better job in predicting hourly and peak ozone concentrations than in predicting ozone precursor concentrations. Such conclusions are limited by several factors, such as the sparseness of available precursor measurements and larger incommensurability issues between model results and observations. For example, NMHC measurements are sparse, both spatially and temporally, and polar oxygenated VOCs were mostly not measured, which limits representativeness of the data. In addition, VOC species in CMAQ are lumped according to functional groups and reactivities instead of carbon numbers. Aggregation of NMHC species from the model to compare with observations introduces some additional uncertainties.

The normalized biases indicate that the model underpredicts NOx and CO, which are mainly emitted from mobile sources, and overpredicts VOCs, which, at the measurement locations, are mainly emitted from area sources. Errors in simulated winds and ABL depth should affect the primary species (NOx, VOC, and CO) similarly, as discussed in Section 3.3.1. Therefore, different signs seen in prediction biases of precursor concentrations may suggest uncertainties in the emission inputs, i.e., an overestimation of area emissions and an underestimation of motor vehicle emissions). That is, these biases can reveal information about the quality of the emissions inventory.

4.3 Conclusions

In this study we have conducted high-resolution photochemical air quality modeling with the MM5-CMAQ for Central California over a four-month period in summer 2000. Initial diagnostic simulations were conducted on a smaller inner domain, and final simulations were conducted for a larger domain after including model improvements resulting from the initial round of diagnostic simulations. We evaluated the temporal and spatial behavior of ozone predictions. We highlighted model performance in predicting ozone and its precursors in conjunction to meteorology evaluation across synoptically driven high-ozone events (exemplified by the four IOPs), and compared them with model performance averaged over the whole summer. By doing so, we have identified features in model performance associated with simulation of different types of ozone-conducive conditions.

The time series of domain-wide aggregated performance metrics indicates that the model has stable temporal performance for the entire summer season. Hourly ozone and 1-h and 8-h peak ozone predictions of each summer day generally have normalized biases within ±15 percent and normalized gross errors less than 25 percent. The model on average underpredicts ozone (hourly or peak) levels, which is also found in previous studies (Tong and Mauzerall 2006; Eder et al. 2006 and 2009) with pair-wise comparison. The best predicted range in 8-h peak ozone is 60~80 ppb. Model performance varies more spatially than temporally. Ozone predictions in the SJV, where highest ozone levels are found, are consistently better than other air basins measured by various metrics. Despite the spatial differences, across all the air basins, the summer average metrics for ozone (hourly or peak) only showed modest underprediction within 10 percent and normalized gross errors around 20 percent. Underprediction of 1-h peak ozone is associated partly with problems in peak timing in the model during the transition hours (5–7 p.m.), which are subject to more meteorological uncertainties than other times of day. Eight-hour peak ozone predictions are consistently better than the 1-h peak ozone.

A rich collection of field measurements enabled this study to evaluate both meteorological and chemical simulations during four synoptically driven high-ozone events. Domain-wide aggregated metrics for predicted ozone and precursor species suggest a consistent level of agreement with observations across all the IOPs, and model performance is similar to the summer average. Episode-to-episode differences in ozone predictions are more pronounced at a subregional level. Summer averages are predicted better than some of the episodes in the SFB and SV. The model exhibits the most robust performance in the SJV, and episodic ozone predictions there are similar to the summer average. This can be attributed, in part, to more accurate predictions of meteorological fields by MM5 for the SJV across all the episodes. Model performance is less satisfactory (normalized peak ozone biases outside ± 15 percent range) in the Sacramento Valley and the Bay area, and is most noticeable in episodes that are subject to the largest uncertainties in meteorological fields. The Sacramento Valley exhibits the poorest agreement in wind directions, which affects the northward transport of pollutants from the metropolitan areas and results in underpredictions of ozone in the northern part of the SV. The strength and direction of the simulated incoming marine flow in the San Francisco Bay area is also subject to more uncertainties, especially during the most stagnant IOP (IOP4). Simulated

ozone downwind of the SFB along the coast increases with the strength of the northwesterly component of the prevailing flow, which explains the overprediction in IOPs 3 and 4 in this region. IOPs 3 and 4 are driven by high-pressure systems with ozone-conducive conditions (e.g., high temperature, clear skies, stagnations). These conditions are generally found to produce satisfactory model performances (e.g., Eder et al. 2006). With a focus on high-ozone events, this study may indicate that in Central California, the model performance is more geographically reliable because accurate simulations of wind directions in the complex terrain of the SV and strength and timing of the onshore flow of the SFB are especially challenging to achieve. Currently, wind profilers deployed in the Central California domain have more spatial coverage in the SJV than other air basins. More meteorological measurements are needed to improve the understanding and characterization of wind flow and other boundary layer dynamics in the SV and SFB, which will subsequently benefit the ozone modeling.

In addition to uncertainties seen in the meteorological fields, evaluation of ozone precursors (the model overpredicts NMHC and underpredicts NOy and CO) points to uncertainties in the emission inventory. The agreement between predicted and measured precursor concentrations is less satisfactory than ozone, and these can be attributed, in part, to sparseness in time and space and larger uncertainties in the measurements. As ozone is a secondary species, precursor observations are invaluable for evaluating photochemical modeling. To improve the understanding of precursor predictions, more real-time measurements of VOC species, improved accuracy of NOy measurements, and increased spatial coverage are needed.

In regions with complex terrain, air pollution events tend to be tied to local geography. Our study shows location-specific differences in model performance under different episodic conditions, which are not revealed when performance metrics are aggregated over the whole domain. Predicted ozone levels in the Sacramento Valley, the San Francisco Bay area, and other coastal regions can be significantly affected by uncertainties in meteorological fields, and as a result their performance metrics differ from the summer average values. Given that ozone responses to emission changes interact strongly with meteorology, as shown in our sensitivity simulations, the effectiveness of emission-control strategies for different episodes or time periods can differ greatly, and this is a limitation of relying on episode-driven simulations for air quality planning.

CHAPTER 5:
Clustering Regional Ozone Concentrations to Reveal Meteorological Regimes Influencing Air Quality[9]

5.1 Introduction

California's Central Valley suffers from serious ozone air pollution problems due to its unique geography and diverse emission sources from both local and upwind areas. The primary flows in the region are produced by the thermal contrast between the ocean and land, and between the valley and surrounding mountains. On typical summer days, westerly winds are funneled into the valley through gaps in the coastal range, along with pollutants transported from the San Francisco Bay area. During the day, the flow is directed up the Sierra Nevada Mountain slope, while at night it reverses and recirculates local pollutants. Meteorological factors can either directly alter ozone chemistry and transport, or indirectly affect ozone formation by influencing light and temperature-sensitive precursor emissions from biogenic sources (Dawson et al. 2007; Tao et al. 2007). As previous modeling studies have often focused on ozone episodes that last for three to five days, knowledge about meteorological regimes that influence ozone behavior has been obtained mainly from statistical analysis of historical observations at limited measurement sites (e.g., Ludwig et al. 1995; Beaver and Palazoglu 2006b).

Investigators participating in the Central California Ozone Study (CCOS) conducted cluster analysis of past ozone data (126 sites, from 1996 to 1998) based on the spatial patterns of several very high ozone days selected by the local air quality districts (Fujita et al. 1999). Three clusters were identified:

1. The SFB cluster has its highest basin-wide ozone when the sea breeze is the weakest, and the SJV ozone is very high during these periods as well.
2. The SJV cluster has its highest basin-wide ozone when the sea breezes are stronger, moving pollutants from SFB and SV to SJV. The SFB and SV clusters are relatively cleaner during these periods.
3. The SV cluster has its highest basin-wide ozone values when there are stronger sea breezes and higher temperatures in the Sacramento Valley. Ozone for the SJV cluster is also high under these conditions.

No distinction was made for ozone patterns within each subregion (San Francisco Bay area, San Joaquin Valley, and Sacramento Valley), although such distinctions may be related to different meteorological regimes.

Cluster analysis is an unsupervised classification method for grouping large amounts of data into smaller categories so that the data in each subset (ideally) share some common trait (Everitt 1993). Observational studies have adopted this method to identify different meteorological

[9] This section is reproduced in part from Jin et al. 2011. J

regimes characterized by distinct associations between air quality and meteorology (Thompson et al. 2001; Schlink et al. 2003), since the approach was initially endorsed by Crutcher et al. (1986). A recent paper by Beaver and Palazoglu (2006a) provides a good review of this method and its application to air quality studies.

Yarnal (1993) identified two fundamental classification approaches to determine various meteorological regimes that affect air quality. The most widely used approach is to first cluster meteorological regimes based on the variability of selected meteorological variables, which are then related back to observed pollutant concentrations (e.g., Fernau and Samson 1990; Eder et al. 1994; Kaufmann and Weber 1996; Cohn et al. 2001; Darby 2005; Beaver and Palazoglu 2006b; Galvez 2007; de Foy et al. 2008). Meteorological regimes derived from this approach depend on the meteorological variables selected, whose variability may not be of influence for pollutant formation and/or accumulation. Classification of a complete set of meteorological variables is technically formidable, because it involves large collections of scalars and vectors that are often different on the most fundamental level (Ludwig et al. 1995). The second approach is to first identify patterns of ozone concentration and then to determine the associated meteorological conditions (e.g., Ludwig et al. 1995; Schwarzhoff and Reid 2000; Beaver and Palazoglu 2006a). This approach is valid only when the changes in ozone concentration patterns are not significantly influenced by factors other than meteorology (e.g., large wild fire emissions). Both approaches yield useful information. Past studies applied these methods to observational data that generally have extensive temporal coverage to account for uncertainties resulting from inter-annual variability and emission changes (e.g., day-of-week differences). However, whether the limited spatial coverage of measurement locations can adequately capture regional level ozone/meteorological patterns has largely been ignored in previous studies.

In the application of these two approaches to a modeling study of ozone formation, sources of the limitations in observational studies can be controlled. In the first approach, the most suitable meteorological variables that are most conducive to ozone formation can be selected as a result of sensitivity analyses. In the second approach, factors other than meteorology such as emission inputs can be specified for simulations of ozone formation. There are no spatial gaps in the gridded model data, and therefore spatial representativeness is no longer an issue. Despite these advantages, few analyses of photochemical modeling results have been guided by this approach. This is because model development and input data preparation generally limit highly resolved simulations to short modeling periods of several days, where temporally representative patterns are difficult to discern. Our study simulates ozone formation for the whole summer of 2000, and thus provides an opportunity to utilize and evaluate the second approach.

This chapter focuses on identifying regimes of modeled ozone spatial behavior in the San Joaquin Valley, and investigates associated meteorological features to reveal the mechanisms for regional ozone transport and accumulation in Central California. This modeling study and data analysis are performed with much higher spatial coverage, and as a result, can better reveal the regional-scale differences between meteorological regimes. Regional-scale differences are mostly influenced by synoptic motions, while measurement locations are affected by finer-scale variations induced by mesoscale flows. Our study can be used to investigate whether

results from historical data are affected by limited measurement sites. Comparing the clustering results from modeling with the ones using historical data also provides information regarding model performance about whether the model is able to capture important ozone spatial distributions.

The ozone clusters and associated meteorological conditions identified from model simulations of the entire summer season can serve as a basis for further diagnostic analyses, such as sensitivity analysis for ozone responses to emission reductions under different ozone regimes, and inter- versus intra-basin transport of pollutants under different meteorological conditions.

5.2 Methods

Cluster analysis and principal component analysis are employed in this study to determine and interpret spatial patterns of ozone in the Central California.

5.2.1 Principal Component Analysis

Both ozone and meteorological fields consist of large spatial dimensions for any given time; for example, 104 dimensions are needed to describe an ozone pattern on one day for a domain of size 100 by 100. These dimensions are redundant and can be reduced due to spatial correlation. To accomplish this, Principal Component Analysis (PCA) has been frequently employed by linearly combining the original large number of dimensions into a few new dimensions without compromising much of the data variance. Reduced dimensionality makes the data set much easier to manage and interpret. This method was first described by Pearson (1902) and Hotelling (1935), and was introduced into meteorology by Lorenz (1956). Since then, it has been widely known in climate science as the Empirical Orthogonal Function (EOF) technique. The basic idea is that the spatial data at any given time can be represented by the sum of the mean pattern and multiples of a few principal spatial patterns (PCs or EOFs). These principal component multipliers (defined as "scores") can be used as a reduced set of descriptors to describe the spatial patterns. The resulting PCs are associated with distinct physical processes when they strongly influence the spatial variations. In this case the scores can be interpreted as indicating the strength of these physical processes. Convenient as the PCA method is, Dommenget and Lafit (2002) stressed that caution should be used when trying to assign significant physical meanings to these statistically derived principal components.

The spatial data in this study include both scalar quantities (e.g., ozone concentrations, surface temperature) and vector quantities (e.g., wind vectors). For scalar quantities, such as 8-h ozone maxima for a specific region, the data matrix (M) is formed as follows:

$$M = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1P} \\ X_{21} & X_{22} & \cdots & X_{2P} \\ \vdots & \vdots & \vdots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NP} \end{pmatrix} \quad (5\text{-}1)$$

where, X is the maximum 8-h ozone concentration on day *i* at grid cell *s*; *N* is the total number of days, and *P* is the total number of grid cells. The anomaly matrix (M') is formed as:

$$M' = \begin{pmatrix} X'_{11} & X'_{12} & \cdots & X'_{1P} \\ X'_{21} & X'_{22} & \cdots & X'_{2P} \\ \vdots & \vdots & \vdots & \vdots \\ X'_{N1} & X'_{N2} & \cdots & X'_{NP} \end{pmatrix} \tag{5-2}$$

where X'_{is} is the departure of X_{is} from the time average at grid cell *s*. The principal components are determined through singular value decomposition (SVD) on M':

$$M' = UDV^T \tag{5-3}$$

where *U* (with dimension $N \times r$, with *r* the rank of M') and *V* (with dimension $P \times r$) are orthogonal matrices, whose column vectors consist of left and right singular vectors, respectively. *D* is a diagonal matrix with singular values *dii*. PCs are thus the column vectors in *V*, and the explained variance by the *ith* PC is $d_{ii}^2 \Big/ \sum_{j=1}^{r} d_{jj}^2$.

Singular value decomposition is used here instead of finding eigenvectors of the covariance matrix[10] of *M* because the temporal coverage (the number of days, *N*) is generally less extensive than the spatial coverage (the number of dimensions, *P*).

The PCs provide a new orthogonal basis for the daily ozone (8-h) maxima. Ozone maxima on day *i*: $\vec{X}_i = (X_{i1}, X_{i2}, \cdots, X_{iP})^T$ can now be represented as follows:

$$\begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{iP} \end{pmatrix} = \vec{\bar{X}} + \sum_{m=1}^{N_{pc}} a_{im} \begin{pmatrix} x_{m1} \\ x_{m2} \\ \vdots \\ x_{mP} \end{pmatrix} \tag{5-4}$$

where, $\vec{\bar{X}} = \frac{1}{N} \sum_{i=1}^{N} \vec{X}_i$ is the average ozone maxima over all *N* days; aim is the score (which is also called the coefficient) of the concentration anomaly $\vec{X}_i{}' = \vec{X}_i - \vec{\bar{X}}$ of day *i* projected onto the *mth* PC: $(x_{m1}, x_{m2}, \cdots, x_{mP})^T$. The PCs in Equation 5-4 are arranged in descending order of importance, as defined by the explained variance. Putting all the days together, scores on the first PC have the largest variability, scores on the second PC are the next, followed by the third, and so on.

[10] We applied PCA to the covariance matrix instead of the correlation matrix used in other studies (e.g., Eder et al. 1994) because the variables are in the same units.

When the spatial fields are vector quantities such as two-component surface winds, a similar kind of analysis can be used to extract the wind vector PCs by forming the data matrix as:

$$M = \begin{pmatrix} u_{11} & v_{11} & u_{12} & v_{12} & \cdots & u_{1P} & v_{1P} \\ u_{21} & v_{21} & u_{22} & v_{22} & \cdots & u_{2P} & v_{2P} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ u_{N1} & v_{N1} & u_{N2} & v_{N2} & \cdots & u_{NP} & v_{NP} \end{pmatrix} \quad (5\text{-}5)$$

where *uis* and *vis* are the wind components at time *i* and grid cell *s*, and PCA follows the same procedure.

5.2.2 Cluster Analysis Methods

Cluster analysis is used in this study to define the spatial patterns of daily 8-h ozone maxima of specified subregions in the CCOS domain. With the clustering algorithm, days that have similar spatial 8-h maximum ozone, with respect to magnitude and spread, are grouped together, and the resulting groups each define a unique ozone pattern.

The dissimilarity measurement (D_{ij}) between any two days *i* and *j* is the Euclidean distance between the 8-hr maximum ozone vectors ($\vec{X}_i$ and $\vec{X}_j$) associated with them.

$$D_{ij} = \sqrt{\sum_{s=1}^{P} (X_{is} - X_{js})^2} \quad (5\text{-}6)$$

where s is the grid index within the subregion, and *P* is the total number of grids. *Xis* and *Xjs* are the 8-hr maximum ozone concentrations at grid s on day *i* and day *j*, respectively.

In most observational studies, the data matrix is scaled to weigh all the measurement locations equally in the analysis; that is, observations from each station are scaled independently by removing their mean over all the days, then dividing by their standard deviation. In contrast, when all the grid cells are included in the analysis as in this study, such scaling is no longer appropriate. Low ozone levels account for the majority of the domain grid cells, while high ozone levels are concentrated near the source regions. Therefore, standardizing each grid cell tends to exaggerate the ozone variations at those places that are always relatively clean. No scaling is applied here so that the ozone patterns are distinguished by their absolute ozone magnitude and spatial extent.

There are two types of cluster analysis methods in general. Hierarchical methods yield an entire hierarchy of clusters of the data set and no re-shuffling can be done within each hierarchy. Once an outlier is assigned to a cluster, it cannot be changed to a more appropriate cluster during the analysis. Partitioning methods, in contrast, divide the data into a predetermined number of clusters, and allow reallocation of data points among the clusters upon each iteration.

In this study, we employ the most widely used partitioning method: K-means clustering (Anderberg 1973). The clustering procedure partitions the ozone days into *k* groups, where the

within-group sum of squares over all the dimensions (here grids) are minimized. For any given *k*, different initial partitioning gives different final clustering results. In order to stabilize the final clustering result, we repeat the process with 50 random starts and pick the one that has the minimum within-group sum of squares (WSS) (See Appendix B) as the final result. This gives us more consistent clustering results. There is no consensus on an objective method for determining the number of groupings, *k* (Everitt et al. 2001). Some studies combined Hierarchical methods with Partitioning methods to aid visual inspections of the data structure for selecting an appropriate number of groupings (e.g., Beaver and Palazoglu 2006a) or used a criterion with single metrics like the Silhouette[11] measure of Kaufman and Rousseeuw (1990). Others were entirely subjective (e.g., Ludwig et al. 1995) so as to have enough clusters to encompass the range patterns in the data set but not too many to greatly reduce the size of each cluster. Although an automated scheme is desirable, as clustering analysis is only a convenient tool for summarizing our data, some expert knowledge can be incorporated *a priori*, as practiced in Darby (2005), to insure that important information is revealed without engaging in over-interpretation at the same time. Increasing *k* for the K-means method generates more compact clusters, and thus generally decreases the within-group sum of squares. To avoid over-interpretation, the difference between WSSk and WSSk+1 are calculated for number of clusters *k* = 1, 2, …, 20. WSSs generally do not decrease significantly after *k* is in the range 4~7. Therefore, $\{4,5,6,7\}$ forms a candidate set for *k*. The final value for *k* is selected after visually examining cluster centers to ensure that important spatial patterns are revealed. Silhouette numbers are also computed for the range of *k* that was considered, and they usually do not differ significantly.

5.2.3 Data and Subregions

As we described earlier, hourly ozone concentrations are simulated with the CMAQ modeling system for the period from 2 June to 26 September 2000. The MM5 (Grell et al. 1994) version 3 was used to prepare the meteorological inputs. Details of model simulation inputs and evaluations against observations can be found in chapters 2 and 4.

Clustering of ozone spatial behavior is conducted for three subregions (Figure 5-1) in Central California that have shown frequent ozone exceedances during the summertime. The three subregions are defined by extending each of the three air basins—SFB, SJV, and SV—into their

---

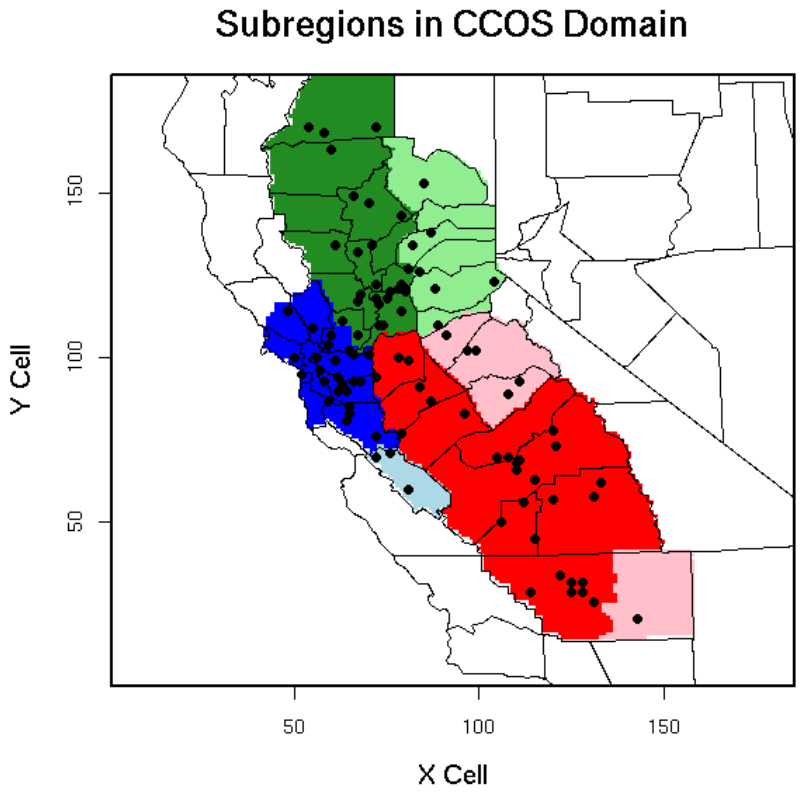
[11] For each observation *i*, the *silhouette number s(i)* is defined as follows:
Let *a(i)* = average dissimilarity between *i* and all other points of the cluster to which *i* belongs (if *i* is the *only* observation in its cluster, *s(i) := 0* without further calculations). For all *other* clusters C, put *d(i,C)* = average dissimilarity of *i* to all observations of C. The smallest of these *d(i,C)* is *b(i)* and can be seen as the dissimilarity between *i* and its "neighbor" cluster, i.e., the nearest one to which it does *not* belong. Finally,

s(i) = (b(i) - a(i)) / max(a(i), b(i)).

Therefore, silhouette number can be used for describing how well a data point fits into the cluster that it is assigned to.

respective downwind areas (indicated by lighter colors), according to the dominant summer flows. Including these additional downwind areas allows us to capture pollutant spatial dispersion patterns that are highly related to the wind regimes.

Figure 5-1: Three Subregions in Central California for Clustering Ozone Spatial Patterns



SJV (red + pink), SFB (blue + lightblue), SV (green + lightgreen). Black dots are actual ozone measurement sites.

Eight-hour ozone maxima are determined for each grid cell for each day. Days with less than 0.1 percent of the grid cells in the CCOS domain exceeding 84 ppb are considered as very clean. These very clean days are labeled as cluster 0, and excluded from further clustering analysis. The Sacramento Valley is heavily influenced by large forest fire emissions from August 18 to 29. These days are also excluded from clustering analysis because their variations are largely induced by emission changes rather than meteorology, as the meteorological changes induced by the fires were not captured by the model.

Weather patterns associated with different ozone clusters were described only with surface meteorological variables in this research, because the prevailing meteorological regimes can significantly affect the ground-level conditions (Beaver and Palazoglu 2009) that are more closely related to ozone formation, accumulation, and transport.

5.3 Results

5.3.1 Ozone Regimes in Central California

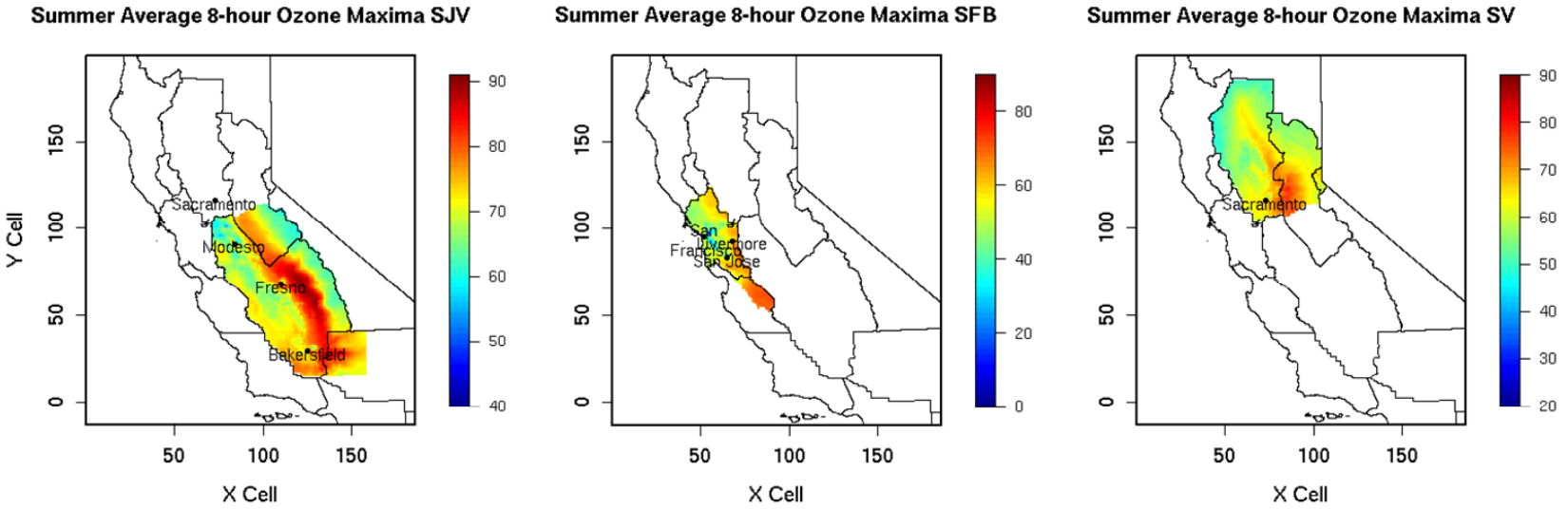
In the first step, PCA was performed for the 8-h ozone maxima for all the days for each subregion, to reduce the number of dimensions needed to describe ozone behavior on each day. Principal component analysis provides insights on underlying physical processes; if one or two processes are particularly influential, they are very likely to be reflected in one of the first few PCs. Figure B2 (Appendix B) reveals that for all the subregions, the first six PCs explain more than 90 percent of the variance in daily ozone spatial maps, and the first three PCs usually account for about 85 percent of the variance. Therefore, a subspace spanned by the first few PCs can represent the spatial ozone anomalies quite well.

Figure 5-2 presents the mean 8-h ozone maxima over the summer. The mean maps indicate ozone hot spots in each subregion, as described in earlier studies (Jin et al. 2008, 2010). On average, the SJV has the highest ozone levels, as indicated by the mean ozone fields.

Figure 5-3 shows ozone loadings on the first three PCs for the three subregions. The first three PCs characterize the most prominent patterns observed in the remaining ozone anomaly fields (deviation of daily ozone maxima from the mean field). PC1 in each subregion is mostly positive, reveals an average spatial trend in ozone concentrations, with higher values downwind of large emission sources and lower values in remote rural areas. PC1 in each subregion tends to explain most of the variability in daily ozone maxima, and this implies that the meteorologically induced ozone changes tend to be positively related to the proximity to anthropogenic emissions. PC2 (and PC3) in each subregion is usually a spatial contrast[12] or gradient in specific directions, characterized by the transition from positive loadings in one part of the region to negative loadings in the other part. For example, PC2 for the SJV is a spatial contrast between the northwestern and southeastern sides of the valley, and larger scores on this PC imply larger differences (stronger contrast) in ozone anomalies between the two sides of the valley. These spatial contrasts in ozone concentrations could be caused by meteorological factors such as a temperature gradient under certain synoptic conditions, or changes in the strength of wind flows that divert pollutants from source regions to downwind areas. In the case of the SJV PC2, for example, the spatial contrast follows the direction of typical up-valley flows, moving pollutants from the source regions in the valley to the Sierra Nevada and the Mojave Desert, which causes a decrease in ozone levels in the northwestern side of the valley and an increase in the southeastern side. A summary description of the first three PCs for each of the subregions can be found in Table 5-1.

[12] *Contrast* is a statistical term usually used in experimental designs. The coefficient (or score, here) of a contrast measures the average difference between treatments that are marked by different signs in the contrast.

Figure 5-2: Mean 8-h Ozone Maxima (ppb) Maps for the Three Subregions



Daily 8-h maximum ozone concentrations are averaged over the whole summer (2 June to 27 Sept.) for the SJV and the SFB, while only the days that are not influenced by forest fire emissions are used to derive the mean map for the SV.

Figure 5-3: Ozone Loadings (ppb) on the Three PCs of 8-h Ozone Maxima

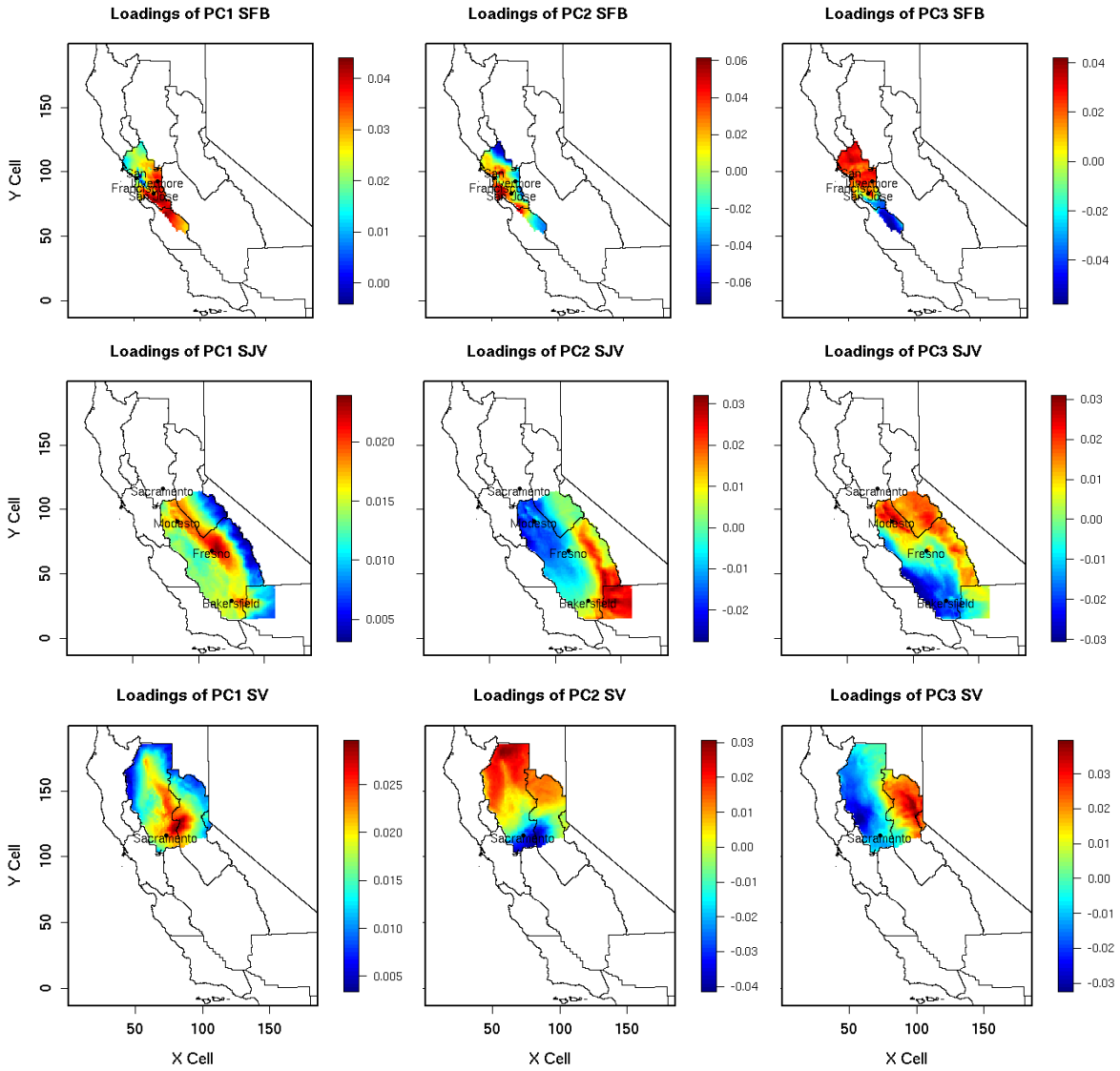


Table 5-1: Description of the First Three PCs for Each Subregion

Subregion	Principal Component	Descriptions
SJV	1	Largely positive, describing a general spatial pattern in ozone concentrations (larger loadings seen in Highway 99, Fresno, and Bakersfield).
	2	A spatial contrast in southeastern direction, with positive loadings in the Sierra Nevada and Mojave Desert, and negative loadings in the northwest (Modesto to Fresno).
	3	A spatial contrast in northeastern direction, with positive loadings on the Sierra Nevada and northeastern SJV, and negative loadings on the southwestern side of the valley.
SFB	1	Largely positive, describing a general spatial pattern in ozone concentrations (larger loadings seen in Livermore valley and south bay).
	2	An east-west spatial contrast between coastal areas and more inland areas, with positive loadings on the coastal side.
	3	A north-south spatial contrast with positive loadings on the northern side.
SV	1	Largely positive, describing a general spatial pattern in ozone concentrations (larger loadings near Sacramento metropolitan area).
	2	A north-south spatial contrast with positive loadings on the northern more rural areas
	3	An east-west spatial contrast between the Sacramento valley air basin and Mountain Counties air basin, with positive loadings in the Sierra Nevada.

Clustering analysis can be performed with the original ozone data, or with the reduced data set by projecting the ozone anomalies onto the subspace defined by first six PCs that generally capture more than 90 percent of the variability in the original data set. Each method produces the same results. Clusters are ranked according to the average scores on the first PC (usually accounting for ~70 percent of the total variances), which indicates the magnitude of the average ozone anomaly levels.

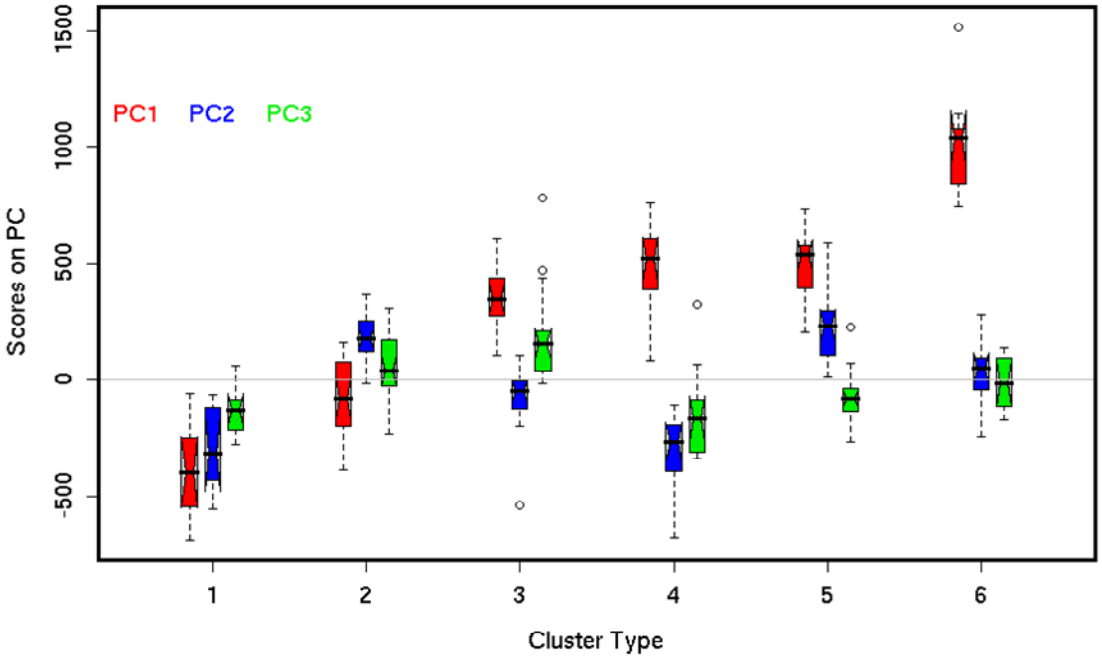
5.3.2 Ozone Regimes in the SJV

The San Joaquin Valley has the most serious ozone problems in this region. After excluding the 27 very clean days (defined in Section 2.3 as cluster 0), 90 days are grouped into six ozone clusters for the SJV. The distribution of cluster members is visualized on the reduced dimensions (Figure 5-4), while mean ozone anomaly fields in each cluster are plotted onto the original dimensions (Figure 5-5). Mean ozone anomaly fields are the cluster mean after the seasonal mean ozone map is subtracted, and they provide a better representation of the spatial features that are unique to the respective clusters. Figure 5-4 summarizes the distribution of the scores on the first three PCs with a boxplot[13] by cluster types to show how each cluster (1 to 6) is significantly different from the others in terms of one or more PC scores.

Clusters 1 and 2 have significantly less ozone than other clusters, as indicated by negative scores on PC1, and they differ from each other by spatial contrasts mainly described by PC2. The spatial contrasts are also reflected in the mean ozone anomaly maps of these two clusters: cluster 1 has elevated ozone on the west side of the valley, while cluster 2 on the east side. **Clusters 3 to 5** have moderate ozone levels with similar PC1 scores, higher than clusters 1 and 2, and lower than cluster 6. These three clusters differ from each other by spatial distributions of ozone anomalies governed by PC2 and PC3. Cluster 3 has especially higher scores on PC3 than clusters 4 and 5. Clusters 4 and 5 differ by having PC2 scores of opposite signs. **Cluster 6** has significantly higher ozone than others (highest PC1 scores). These features are in accord with spatial patterns of mean ozone anomalies (Figure 5-5): Cluster 3 has elevated ozone levels on the north side, cluster 4 on the western side, and cluster 5 on the more downwind southeastern side.

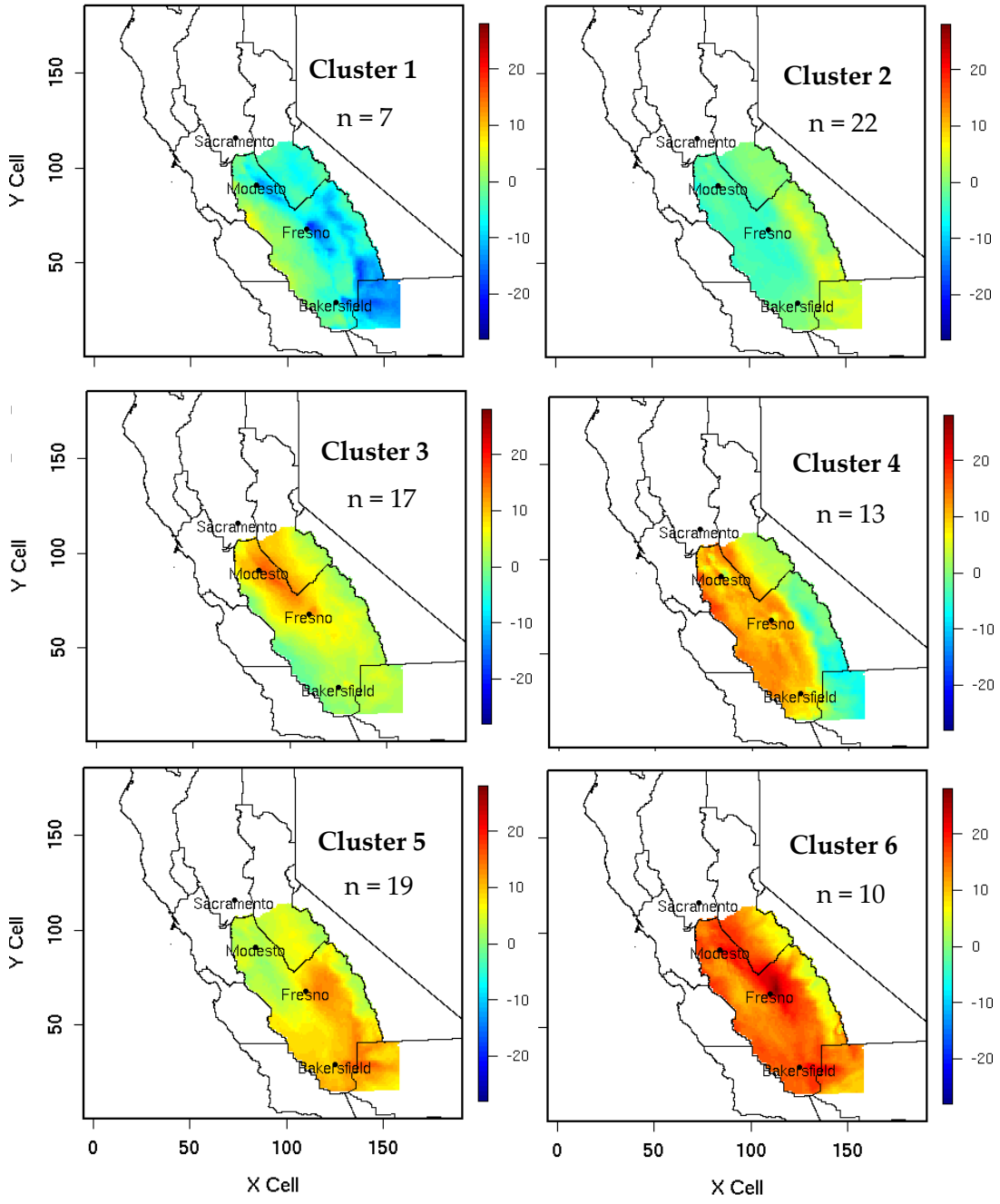
[13] The boxplot produces a box and whisker plot for scores on each PC. The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers. If there are no data outside the whisker, a dot is placed at the bottom whisker. In a notched box plot, the notches represent a robust estimate of the uncertainty about the medians for box-to-box comparison. Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5 percent significance level.

Figure 5-4: Visualizing SJV Ozone Clusters (1 to 6) with PC Scores



Notched boxplot of scores on first three PCs by SJV ozone clusters, showing how each cluster is statistically different from others.

Figure 5-5: Ozone Anomaly Field (ppb) Patterns Averaged over SJV Clusters



(Mean ozone anomaly pattern for cluster *i*) = (Mean ozone pattern for cluster *i*) – (Mean ozone pattern over all days).

The dynamic relationships among the SJV clusters can be revealed by modeling the transitional probabilities (e.g., Beaver and Palazoglu 2008); however, their statistical significance cannot be determined due to the limited number of days simulated in this study. The number of days

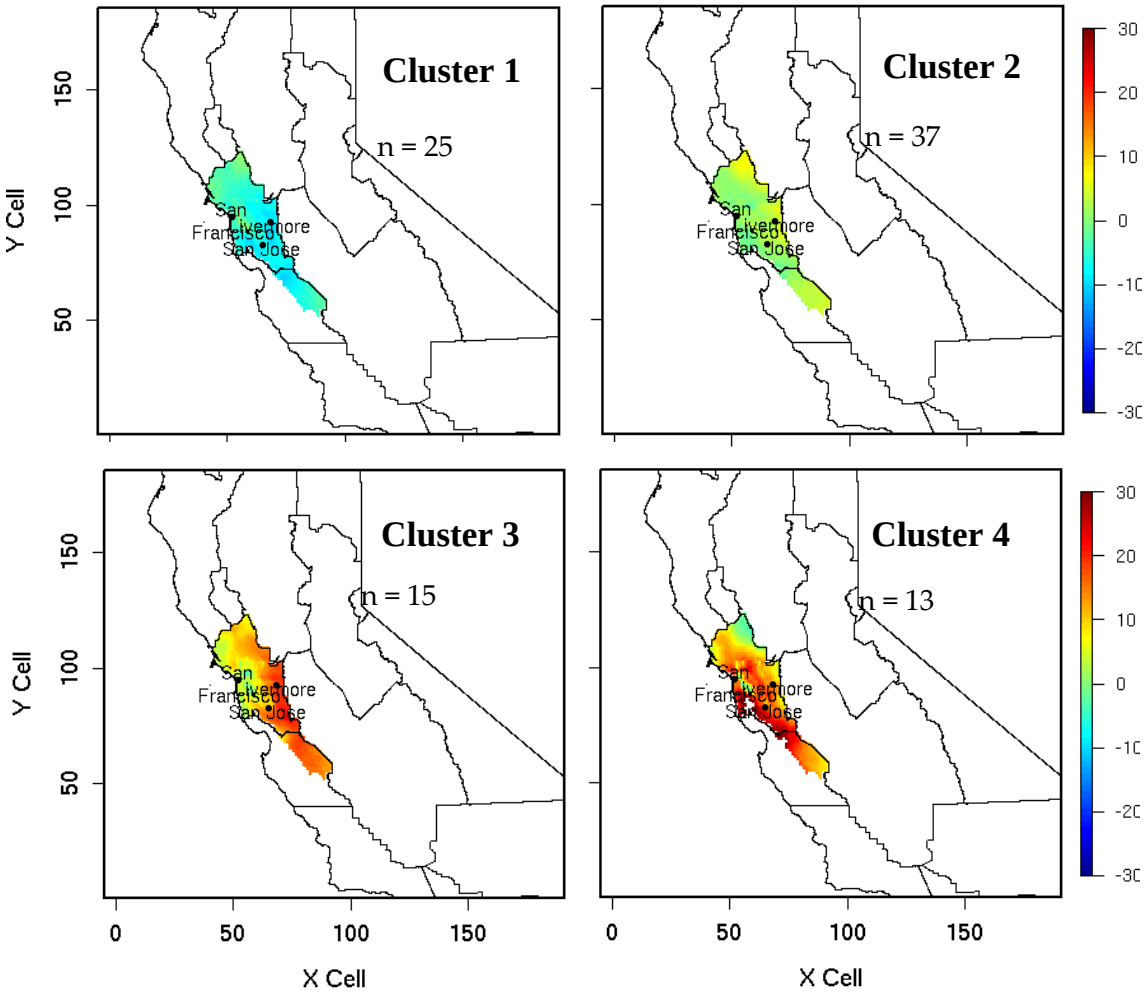
when one cluster made a transition to another (See Supporting information Table B1 in Appendix B) indicates that low-ozone days are less likely to develop into high-ozone clusters on the next day. This implies that ozone accumulation in the SJV usually takes multiple days.

5.3.3 Ozone Regimes in the SFB and the SV

The Bay area and the Sacramento Valley are at times influential upwind source areas for the San Joaquin Valley. The remaining 90 days (after clean days are excluded) are grouped into four SFB ozone clusters. Five clusters are obtained for the SV after 12 additional days, which were significantly affected by forest fires, are excluded.

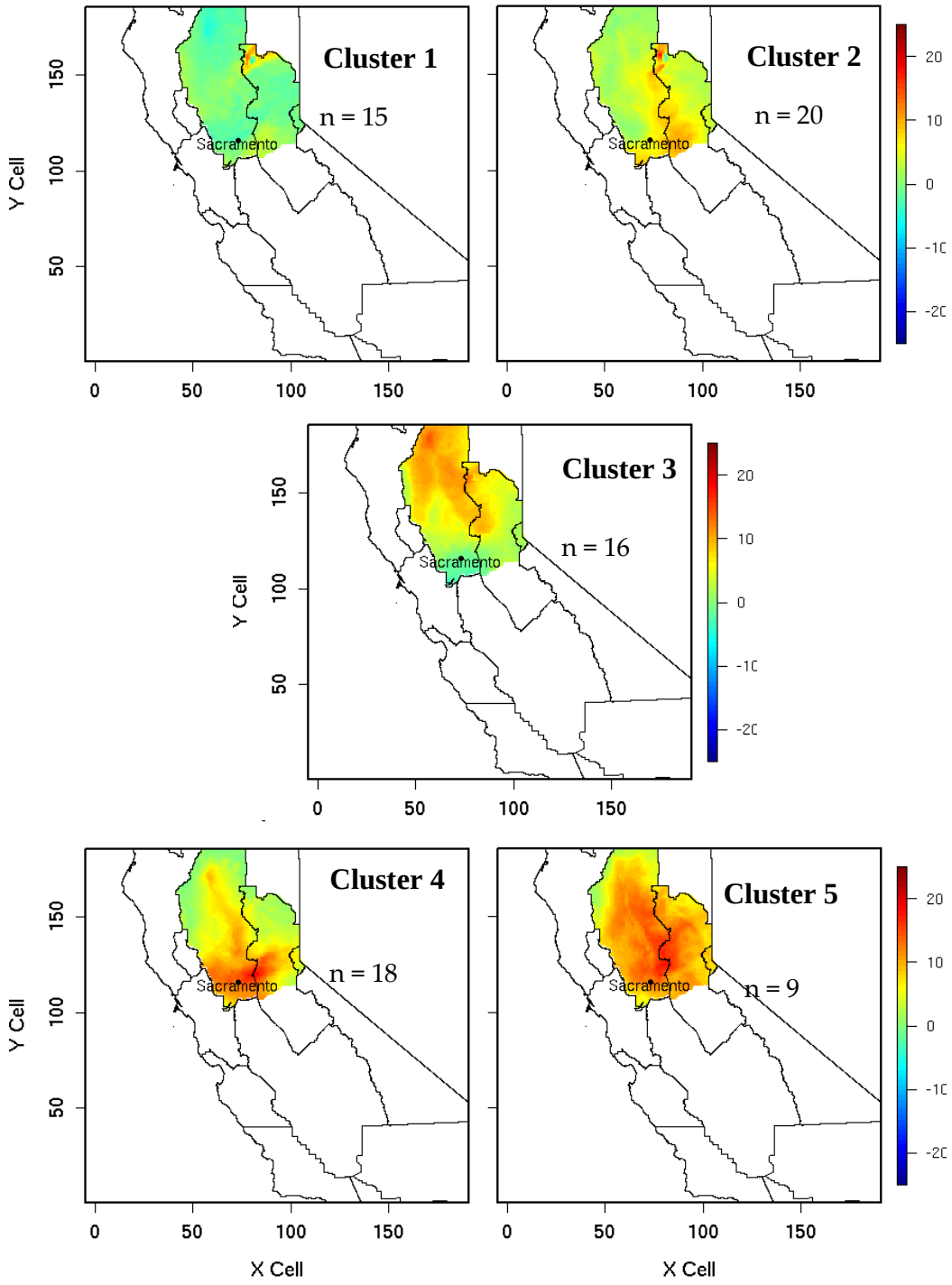
The average spatial patterns of ozone anomalies indicated in figures 5-6 and 5-7 are in accord with the scores on PCs of SFB and SV (Figure B3 in Appendix B). SFB cluster 4 has higher ozone levels near the ocean (western and southern) than it has in the more inland areas (in the southeastern SJV), as in clusters 1 to 3. This is likely due to a decrease in strength of the sea breeze. SV cluster 3 experiences an ozone decrease in the Sacramento urban areas and an increase in the northern downwind side. With ozone ventilated to the downwind areas, locations near Sacramento become rather clean as in SV cluster 3, despite moderately high PC1 scores. SV cluster 5 is characterized by overall high ozone levels both in the Sacramento area (thus high PC1 scores in Figure B3 in Appendix B) and the downwind, northern part of the valley (thus high PC2 scores in Figure B3 in Appendix B).

Figure 5-6: Mean Ozone Anomaly Fields (ppb) Patterns of SFB Clusters



(Mean ozone anomaly pattern for cluster *i*) = (Mean ozone pattern for cluster *i*) – (Mean ozone pattern over all days).

Figure 5-7: Mean Ozone Anomaly Fields (ppb) Patterns of SV Clusters



The characteristics of each of the clusters are summarized in Table 5-2 along with the SJV clusters for future reference.

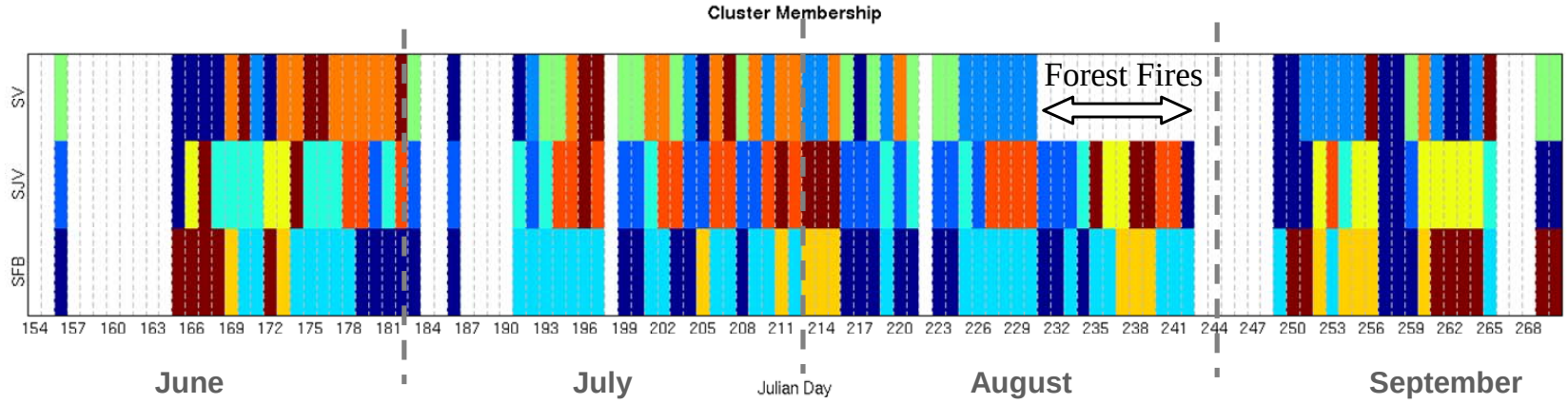
Table 5-2: Descriptions of Ozone Clusters

Subregion	Ozone Level	Spatial Pattern Cluster	Descriptions of Ozone Anomalies
SFB	Low	1	Mostly negative, indicating lower than average ozone levels
		2	Very small ozone increases on the eastern side (more to the inland side) of the basin
	High	3	Elevated ozone concentrated on the southeastern side of the basin
		4	Elevated ozone levels on the western, central, and southern side (more to the ocean side) of the basin
SJV	Low	1	Overall lower than average ozone levels, with very small increases on the western side of the valley
		2	Small ozone increases in the Sierras and the Mojave Desert, downwind side of the major urban areas
	Moderate	3	Moderate ozone increases on the north side
		4	Moderate ozone increases on the western side
		5	Moderate increases on the southeastern
	High	6	Strong ozone increases throughout the valley
SV	Low	1	Very small ozone increases in the Sacramento area (southern end of the valley)
		2	Small ozone increases near Sacramento
		3	Strong ozone increases in the north, but decrease in the Sacramento, resulting in rather clean conditions in the usual ozone hot spots
	Moderate	4	Moderate ozone increases both in the south and southeast Sacramento area
	High	5	Strong ozone increases throughout the valley

Subregion daily cluster memberships are plotted for the entire simulation period in Figure 5-8. Similar to the SJV, the highest ozone pattern in the Sacramento valley does not follow the cleanest days (cluster 0, 1, and 3) but rather evolves over multiple days. The SFB is a more

accessible area that is more sensitive to mesoscale meteorology (e.g., changes in sea breeze intensities) by adjusting its ozone levels more quickly. A clean day in SFB (cluster 0, 1, and 2) can be followed by a high ozone day (cluster 3 and 4), which entails an ozone increase by as much as 10 ppb in the ozone hot spots (e.g., Livermore valley, San Jose), according to the mean anomaly ozone maps shown in Figure 5-6.

Figure 5-8: Cluster Memberships



Clean days (cluster 0) are blank. Days with significant forest fire influences on SV ozone levels are indicated on the plot.

SJV Key (cluster type from 1 to 6)

SFB Key (cluster type from 1 to 4)

SV Key (cluster type from 1 to 5)

The joint distribution of ozone clusters in SJV in relation to SFB and SV subregions is presented in Table 5-3. Moderate high ozone clusters 3 and 5 in the SJV are mostly associated with a relatively clean SFB ozone pattern (cluster 2), while SJV cluster 4 is mostly associated with high ozone SFB clusters 3 and 4. In this analysis period, high ozone levels in SFB (cluster type 4) occur when SJV experiences its lowest ozone levels. This is caused by the high ozone sensitivity to SFB meteorology. SFB produces and accumulates ozone over the course of a day when it experiences a sudden weakening of westerly flows. Since the transport of pollutants from the SFB is diminished, the SJV maintains its low ozone levels due to its buffering ability. There are fewer clear relationships between SJV and SV ozone patterns, partly due to smaller sample sizes in the SV. When SV is experiencing its highest ozone levels, SJV ozone levels are generally high (clusters 3–6), and low SJV ozone levels (clusters 1 and 2) do not occur on the highest SV ozone days (cluster 5). Among SJV moderate ozone days (clusters 3, 4, 5), clusters 3 and 5 are associated with higher SV ozone levels than cluster 4.

Table 5-3: Inter-Basin Relationships of Ozone Clusters

SJV cluster type							SJV cluster type						
SFB	1	2	3	4	5	6	SV	1	2	3	4	5	6
cluster	1	2	17	3	0	3	cluster	1	5	3	2	4	0
type	2	2	4	11	1	16	type	2	1	3	4	4	6
	3	0	1	2	6	0		3	2	10	2	0	2
	4	5	0	1	6	0		4	0	3	4	2	6
								5	0	0	4	1	3

Each entry at *ith* row and *jth* column is the number of days that has ozone cluster types *i* and *j* in the respective subregions.

5.3.4 Meteorological Features by Ozone Regimes

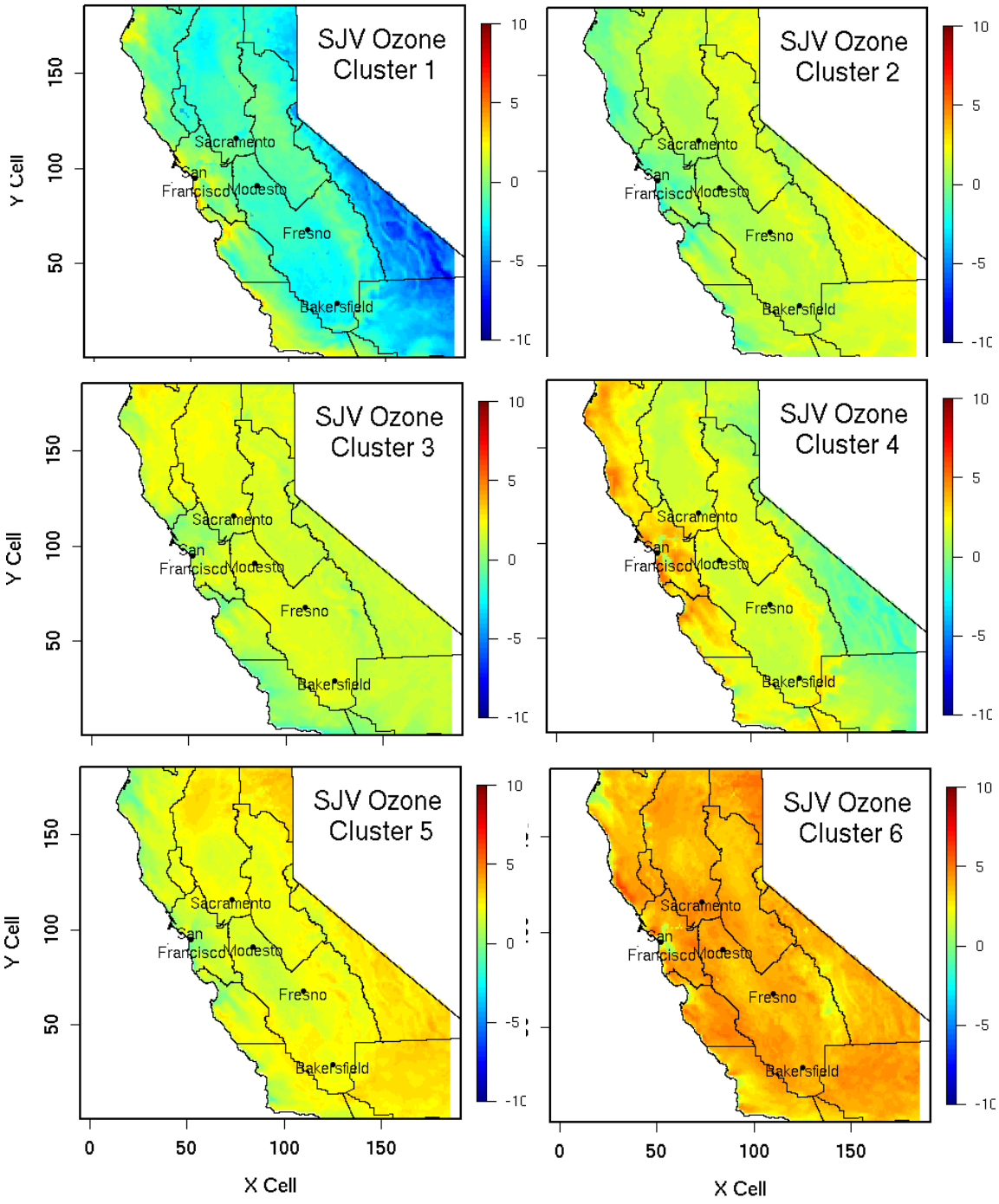
In this section, we focus on the spatial patterns of ozone in the SJV and investigate whether distinct meteorological conditions that explain the ozone transport and accumulation processes are associated with each ozone cluster. Forest fire emissions could confound the effects of meteorological variations on ozone concentrations, but during the period of this analysis, the SJV was not significantly affected by wildfires. The effects of other anthropogenic emission changes such as weekend emission reductions are found to be much smaller than the meteorological effects (see the discussion in Section 5.4). Changes in ozone spatial behavior are hence mainly attributable to changes in meteorology. Two meteorological variables are considered here: daily maximum temperature (8-h running average), and surface winds, which were found previously to be significant in describing the daily maximum ozone concentration with using statistical models (Schlink et al. 2003).

5.3.4.1 Temperature Regimes

Principal component analysis reveals that the first component of the temperature fields accounts for the most variance (~80 percent) associated with temperature changes, while the first three components explain 90 percent of the variance. Component 1 has similar loadings throughout the domain, and thus corresponds to a relatively homogeneous warming or cooling in the area (Figure B4 in Appendix B). PC2 and PC3 both represent temperature gradients between the inland and coast but in different directions.

Mean temperature anomaly fields associated with each SJV ozone cluster are presented in Figure 5-9. The last cluster (with highest ozone levels throughout the valley) has the strongest increase (~4–5 degrees) in temperature anomalies across the whole domain. Clusters 1 and 4 are each characterized by a stronger temperature increase in coastal areas relative to inland, although the overall temperature in cluster 1 is much lower than cluster 4. Similarly, clusters 2 and 5 have a higher temperature increase in areas further inland (Northeast Plateau, Great Basin Valleys, and Mojave Desert) than the coastal regions, although cluster 2 is generally cooler than cluster 5. Cluster 3 also has cooler temperatures in the Bay area and central coast, but higher temperatures in the northern portion (mainly Sacramento Valley) of the domain. Higher temperature anomalies, in general, contribute to higher ozone anomalies. Cluster 3 has higher temperature and higher ozone in the northern part of SJV compared to cluster 5, which has both higher temperature and ozone in the southern part of the valley. A different explanation is required for cluster 4. The highest ozone anomalies in this cluster are located on the western side of the valley, where temperature anomalies are not significantly higher than cluster 3 and 5 (according the mean scores on PC1).

Figure 5-9: Average Temperature Anomalies (K) for Each SJV Ozone Cluster



5.3.4.2 *Wind Regimes*

The flow regimes are also important for pollutant dispersion and accumulation but cannot be represented by single metrics (like daily maximum temperature), because both wind speed and

direction tend to change during the course of the day. Hourly wind fields are examined to determine the diurnal behavior in the flow regimes associated with SJV ozone clusters.

Principal component analysis shows that the first three components of the wind fields account for ~60 percent of the variance (Figure B5 in Appendix B). Unlike other PCA results (ozone and temperature), the distribution of wind variances has a longer tail, and more PCs need to be included to capture 80 percent of the total variance.

Mean wind fields and the first two principal components are shown below with the color coded wind speeds (Figure 5-10).

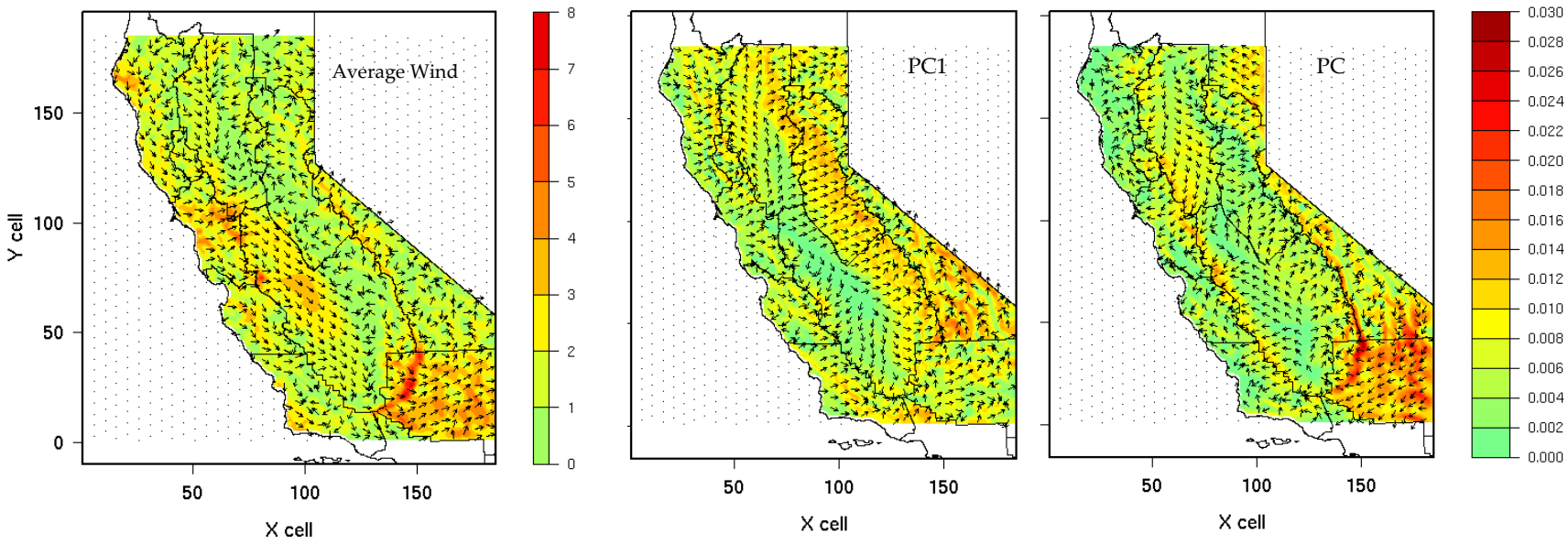
The mean pattern shows a strong westerly flow from the SFB into the SJV that turns southward. Highest wind speeds are found at the southern end of the domain, where wind speed increases over the Mojave Desert. The mean flow reveals a cyclonic eddy to the east of Fresno, and also near Sacramento, where mesoscale phenomena known as the "Fresno eddy" and the "Sacramento eddy" (Lin and Jao 1995) occur. When the wind anomalies are close to zero in these places, these eddies are present in the actual flow patterns (mean + anomaly).

PC1 emphasizes a westerly up-slope ventilation flow toward the Sierra Nevada. When the score is positive, there is an increase in pollutant transport toward the Sierra Nevada from source regions in the SJV, such as Modesto, Fresno, and Bakersfield. When the score is negative, the eastward transport is weakened or even reversed into a down-slope flow.

PC2 emphasizes the reverse flow from the Central Valley toward the Bay area and coastal areas through the gaps (Tracy station, Pacheco Pass, etc.) along the coastal ranges, as well as reverse flow from the Mojave Desert to the southern part of SJV. It also emphasizes the up-slope flows on the eastern side of the Sierra Nevada, which are more or less disconnected from the SJV geographically and have much less effect on ozone levels in the valley. A negative score means an increase in inflow from the SFB and coastal regions and increased dispersal of pollutants from the southern SJV (mainly Bakersfield area) to the Mojave Desert. A positive score implies reduced sea breezes and reduced ventilation in the southern SJV.

PC3, not shown here, emphasizes the north-to-south flow. Wind fields have only ~7 percent variation for this component, implying that changes in wind fields in this dimension are much smaller relative to PC1 and PC2. Wind variations on other PCs are even smaller than those on PC3, and present no obvious physical meanings.

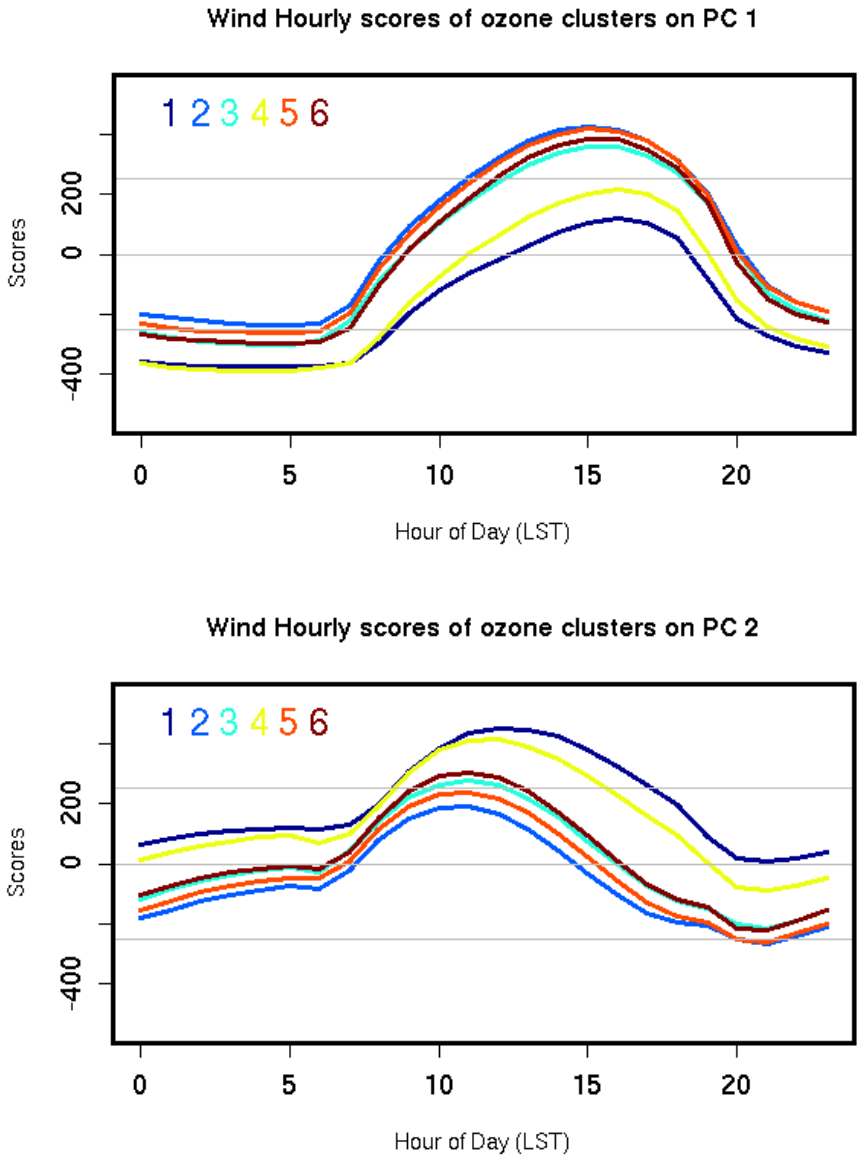
Figure 5-10 Average Wind Fields and First Two PC Loadings



Wind speeds (m/s) are color coded.

The diurnal flow patterns of each ozone cluster are projected onto the first two PCs with mean scores plotted (Figure 5-11). Diurnal profiles of scores on the third principal component among each cluster are not very different, and are not shown here.

Figure 5-11: Diurnal Profiles of Surface Wind Scores



Shown for the first two PCs for SJV ozone clusters.

The scores on PC1 describe diurnal changes in the up-slope flow component in the wind fields. Across the clusters, similar patterns are observed in the diurnal variation: up-slope flow starts to increase (or down-slope flow start to decrease, when the score is negative) around sunrise

(7 a.m.) until it reaches a maximum around 3 p.m., then it starts to decrease and reverses to a down-slope flow at night. The scores on PC2 describe diurnal changes in an off-shore flow component (reversal of sea breeze) on top of the strong sea breeze in the mean wind map (Figure 5-10), represented by the flows through mountain passes along coastal ranges. Diurnal changes in this component are largely driven by the temperature difference between land and sea. One sees that in general, sea breeze begins to increase around 10 a.m. (indicated by a negative slope of the curve), when the land heats up, about three to four hours after sunrise. It reaches maximum speed around sunset at 8 p.m., and slowly decreases at night (small positive slope) and in the early morning before 10 a.m.

Despite similar variations in the diurnal wind profiles, in actual wind speed, weakest daytime up-slope flow is seen in clusters 1 and 4, and strongest in clusters 2 and 5. The difference between the two sets in the maximum up-slope flow speed (around 3 p.m.) is about 2.5 m/s. Duration of this ventilation flow in clusters 1 and 4 is also shorter (from around noon to 6 p.m.) than that of cluster 2 and 5 (from 8 a.m. to 8 p.m.). At night, down-slope flows in clusters 1 and 4 are stronger than those in clusters 2 and 5. With weakened upslope flows in clusters 1 and 4 during the day and stronger downslope flow at night, more pollutants are trapped in the valley, causing more accumulation of ozone along the western side of the valley and less spread to the eastern downwind areas.

Similarly, as indicated by the scores on PC2, clusters 1 and 4 have the weakest sea breezes throughout the day, and wind speeds are about 2.5 m/s less than those of clusters 2 and 5, which have the stronger sea breezes. The reduction in sea breeze provides an indication of stagnant conditions in clusters 1 and 4. With the reduced sea breeze, pollutants transported from the SFB and other coastal regions to the SJV are reduced and more limited to the western side of the valley, which further strengthens the ozone pattern in clusters 1 and 4. At the same time, weaker sea breeze contributes to SFB ozone accumulation. Along with higher temperature anomalies seen along the coasts (Figure 5-9, cluster 4), meteorological conditions in SJV cluster 4 lead to high ozone levels in both the SJV and the SFB, as seen Table 5-3. Outflow from the southern SJV to the Mojave Desert is also weakest in clusters 1 and 4, resulting in lower ozone levels downwind of Bakersfield than those seen in clusters 2 and 5 (Figure 5-5).

Strong sea breezes during the days as seen for clusters 2 and 5 ventilate SFB pollutants downwind, reducing ozone levels in the SFB and northern part of the SJV. The flow patterns captured in clusters 2 and 5 help to explain the inter-basin relationship of ozone patterns in the SFB and SJV indicated in Table 5-3, that SJV clusters 2 and 5 are mostly associated with the SFB cluster 2 with low ozone levels.

Different flow patterns described by PC1 and PC2 explain the unique feature seen in cluster 4, where temperature anomalies alone cannot account for its difference from ozone cluster 5. Flow patterns in clusters 3 and 6 are between the two extremes described above; both are characterized by a stronger sea breeze and upslope flow than those in clusters 1 and 4.

Sea surface pressure is another way to understand the flow patterns. Figure B7 in Appendix B presents the sea surface pressure maps of SJV clusters. The mean sea surface pressure maps

across all clusters generally show the pressure gradient induced by mesoscale thermal contrast between land and ocean as well as between valley and surrounding mountains. In the anomaly maps, we can see more synoptic effects on regional level pressure gradients that induce surface flow anomalies superimposed on the diurnal variations common to all the clusters. In clusters 1 and 4, there is an enhanced off-shore pressure gradient relative to the summer average, which weakens the on-shore westerlies and causes stagnation. In clusters 2 and 5, there is an enhanced on-shore pressure gradient relative to the summer average, which reinforces the on-shore flow from northwest to southeast that ventilates the Bay area and leads to upslope transport from the SJV. Cluster 3 indicates a northward gradient, which may explain the pollutants accumulation in the northern part of the SJV. Cluster 6 exhibits less changes in pressure gradients and the region-wide wind patterns are more similar to the summer average.

A more complete picture of SJV ozone clusters 3 to 6 (moderate to high ozone) can now be described in terms of ozone spatial patterns, upwind ozone levels, and regional meteorological conditions (Table 5-4).

Table 5-4: Characteristics of SJV Moderate to High Ozone Clusters

SJV Cluster 3	Ozone distributions	SJV	Moderate ozone, higher than averge ozone in the northern part
		SFB	Generally low ozone levels (cluster 2)
		SV	Moderate to high ozone levels (clusters 2,4,5)
	Meteorological conditions	Temperature	Moderate temperature, higher than average temperature seen in the northern part of the SJV and the domain (mainly SV)
		Flow Patterns	Moderate sea breeze and upslope flow during the day
SJV Cluster 4	Ozone distributions	SJV	Moderate ozone, higher than average along the western side of the valley
		SFB	Generally high ozone levels (clusters 3 and 4)
		SV	Moderate to low ozone levels (clusters 1,2)
	Meteorological conditions	Temperature	Moderate temperature, higher-than-average temperature seen along the coast
		Flow patterns	Weakest sea breeze and upslope flow during the day; weakest outflow in the southern SJV
SJV Cluster 5	Ozone distributions	SJV	Moderate but higher than average ozone occurs in the southeastern downwind areas (of Fresno and Bakersfield).
		SFB	Generally low ozone levels (cluster 2)
		SV	Moderate ozone levels (clusters 2,4)
	Meteorological conditions	Temperature	Moderate temperature, higher than average seen in the southeastern part of the SJV and in the Sierra Nevada and Mojave Desert
		Flow patterns	Stronger sea breeze and upslope flow during the day
SJV Cluster 6	Ozone distributions	SJV	High ozone throughout the whole SJV
		SFB	Mostly high ozone (cluster 3)
		SV	Moderate ozone levels (clusters 2,4)
	Meteorological conditions	Temperature	Highest temperature throughout the valley
		Flow Patterns	Moderate sea breeze and upslope flow during the day

5.4 Discussion

5.4.1 Effects of Weekend Emission Reductions on Spatial Distribution of Ozone

The variations of ozone patterns induced directly and indirectly by changes in meteorology can be confounded by changes in anthropogenic emissions. When using ozone spatial patterns to reveal meteorological regimes, it is important to understand whether ozone patterns are significantly affected by changes in anthropogenic emissions. Anthropogenic emissions have day-of-week variations, with lower weekend emissions than on weekdays in most cases (see Tonse et al. 2008). To investigate the weekend effect on ozone spatial patterns, we simulated all the weekend days with weekday emissions and examined the variations of ozone spatial patterns induced by the emission change. Ozone differences on Sundays (*ΔO3wnd*) were used in this analysis.

$$\Delta O_3 wnd = O_3 a - O_3 h \qquad (5\text{-}7)$$

where, *O3a* are the Sunday ozone concentrations simulated using actual Sunday emissions, and *O3h* is the Sunday ozone concentrations simulated using typical weekday emissions. Saturdays are excluded from the analysis because they can be affected by carryover of Friday emissions.

ΔO₃wnd is projected onto the same SJV ozone PCs, 1 to 3, described in Section 5.3.1. The scores (shown in Figure 5-12) can be compared to the overall variations in ozone anomalies observed for whole summer (Figure 5-4). Scores in Figure 5-12 suggest that the reduction of weekend emissions will reduce overall ozone levels (PC1) and increase the spatial imbalances (PCs 2 and 3). Compared to Figure 5-4, these changes are small relative to the overall variations seen for the whole summer and within each cluster. For example, in Figure 5-4, the scores on PC1 across all clusters range from -500 to more than 1,000. Within each cluster, the scores on PC1 generally vary by more than 300; while weekend effects only induce average variations of ~100. These minor effects are confirmed by clustering SJV ozone patterns simulated using all weekday emissions. The new results are not significantly different from the ones we presented previously (see Table 5-5) in that, the tabulation showed only a few off-diagonal elements. The off-diagonal elements can be explained by emission changes occurring on weekends. For example, in the last column, the overall decreases in weekend ozone concentrations lead to assigning three days in cluster 6 to cluster 5. This confirms that the variations seen in ozone spatial patterns are mostly attributable to changes in meteorology.

Figure 5-12: SJV Weekend Effects on Ozone Spatial Patterns Projected onto the First Three PCs

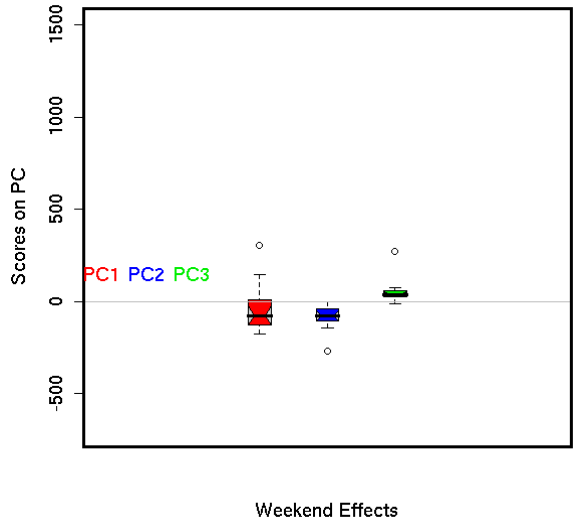


Table 5-4: Comparison of SJV Clustering Results with Constant Emissions and Day-of-Week Emissions[14]

SJV Clusters (c1 ~ c6)	Meteorology Induced Variation
Meteorology + Emission	c1 c2 c3 c4 c5 c6 c1 8 0 0 0 0 0 c2 0 15 3 0 4 0 c3 0 1 14 1 0 0 c4 1 0 0 12 0 0 c5 0 0 4 0 12 3 c6 0 0 1 0 0 9

5.4.2 Representativeness of Ozone Observational Sites

A historical view of Central California ozone patterns was developed from similar analysis using observed data at surface measurement stations. Fujita et al. (1999) analyzed ozone observations from 1996 to 1998 at 126 sites in the CCOS domain to determine spatial patterns of

[14] An outlier day 265 is removed from the clustering due to its exceedingly large score on PC 3.

high ozone days. A comparison of model-simulated ozone patterns and associated meteorological conditions in summer 2000 with those seen in observations provides an opportunity to determine (1) whether summer 2000 is typical of other years, and (2) whether the model is predicting spatial distributions of ozone for the right reasons.

Before comparing our clustering results with those from historical data, the following questions need to be assessed: (1) are there adequate number of observational sites, and (2) are they appropriately located, to capture region-wide ozone patterns as effectively as those in the highly resolved model grids. The observational sites are plotted in Figure 5-1 for each of the three subregions.

The same clustering procedure that was used with all the subregion grids is now repeated using modeled data but only at observational sites. Clean days (cluster label 0) here are defined to be those with no more than one site exceeding 84 ppb (8-h maximum ozone). Resulting cluster memberships are compared with those from modeled data in Table 5-6.

Table 5-5: Comparison of Clustering Results with Modeled Data at All the Grids vs. Using Modeled Data Only at Observational Sites

SJV

	sites						
grids	**0**	**1**	**2**	**3**	**4**	**5**	**6**
0	*25*	2	0	0	0	0	0
1	3	*4*	0	2	0	0	0
2	1	9	*8*	4	0	0	0
3	0	0	0	*11*	1	5	0
4	0	0	0	1	*3*	8	1
5	0	0	2	0	0	*17*	0
6	0	0	0	0	1	1	*8*

SFB

	sites				
grids	**0**	**1**	**2**	**3**	**4**
0	*25*	0	2	0	0
1	3	*15*	7	0	0
2	1	2	*32*	2	0
3	0	0	1	*12*	2
4	0	0	0	0	*13*

SV

	sites					
grids	**0**	**1**	**2**	**3**	**4**	**5**
0	*24*	0	1	1	0	0
1	4	*10*	1	0	0	0
2	0	4	*16*	0	0	0
3	0	1	1	*14*	0	0
4	0	0	2	0	*12*	4
5	0	0	0	1	0	*8*

The bold-faced numbers are cluster labels (row labels are clusters resulting from using modeled data at all grids, and column labels are clusters resulting from using modeled data at measurement sites).

One can see that cluster memberships determined by the two approaches are in reasonable agreement for both the SFB and the SV, since there are few off-diagonal elements. Measurement sites in SJV can distinguish relatively high ozone days (clusters 3, 4, 5, 6) from relatively low ozone days (clusters 1 and 2) and clean days (cluster 0). As for more subtle differences in the clusters, however, SJV observation sites could not quite distinguish cluster 1 from cluster 2 or cluster 4 from 5 (highlighted in red). About 70 percent of the days that should be in cluster 4 have been grouped into cluster 5. These clusters are separated by the spatial ozone contrast

between the western and eastern sides of the SJV, which may be caused by different meteorological conditions that yield different ozone production and transport patterns. Figure 5-1 indicates measurement sites in the SJV tend to be located in more populated areas in the middle or eastern side of the valley. The western side of the SJV is under-sampled. As a result, the current distribution of measurement sites in the SJV alone cannot resolve the spatial contrast between the western and eastern sides of the valley adequately, so that days with spatial pattern of cluster 4 and 5 tend to be lumped together. Recall that SJV cluster 4 is generally associated with high SFB ozone, while SJV cluster 5 is linked to low SFB ozone. Including SFB sites in the clustering procedure used by Fujita et al. (1999), can help to distinguish days in SJV ozone cluster 4 from those in SJV ozone cluster 5 without knowing the actual ozone concentrations on the western side of the SJV.

According to the above analysis, ozone patterns are reasonably well represented by existing measurement sites in the SFB and the SV. SJV measurement sites alone can separate high ozone days from low ozone days but cannot characterize concentration changes on the western side of the valley. By pooling all the measurement sites into the clustering procedure as used for grouping the historical high ozone days, days with different SJV ozone patterns can still be separated. Clustering results derived from historical data can be compared with those from model results to examine whether summer 2000 conditions are representative of the other years.

5.4.3 Comparison with Historical Data

Fujita et al. (1999) used observed ozone from 1996 to 1998 in the entire CCOS domain to determine spatial patterns of "very high ozone days" (moderate to high in this study). Three clusters were identified: (1) the SFB has its highest ozone, while SJV ozone is also very high, which is characterized by the weakest sea breeze; (2) the SJV has its highest basin-wide ozone, while SFB and SV are relatively clean, which is characterized by a stronger sea breeze, moving pollutants from SFB and SV to SJV; (3) SV is at its highest ozone values, which, similar with the second cluster, has a stronger sea breeze but higher temperature in the Sacramento Valley; SJV ozone in this cluster is also high. Cluster 1 occurs most frequently in the historical data set (22 times), followed by cluster 2 (12 times), and then cluster 3 (8 times).

As described previously, data collected at these sites are able to separate days with different SJV ozone spatial patterns that are derived with ozone data with regional coverage. Consequently, the three ozone clusters described above can be compared with modeled SJV ozone clusters to determine whether modeled ozone in summer 2000 captures the important regimes seen historically.

The characteristics of the three historical ozone clusters are reflected in our model results by the SJV ozone clusters 4, 5, 3, their associated meteorological conditions, and the relationship to upwind ozone levels, as shown in Table 5-4. This provides an indication that the model is able to capture ozone patterns for the right reasons. Highest SJV ozone levels appear in cluster 6 in summer 2000, characterized by especially high temperatures throughout the domain and somewhat less stagnant conditions than cluster 4 (which was the most frequent cluster in previous years). Cluster 6 is generally developed from ozone clusters 4 and 5 other than itself,

and is largely associated with high SFB ozone levels and moderate SV ozone levels, and thus is not typical of any of the historical cluster types.

Figure 5-8 indicates that these historical ozone clusters appear in summer 2000 episodes: **days 261 to 264** (September 17 to 20) are characterized by SJV cluster 4, high ozone in the SFB and moderate to low ozone in the SV; **days 227 to 230** (August 14 to 17) are characterized by SJV cluster 5, low ozone in the SFB and moderate low ozone in the SV; and **days 175 to 177** (June 23 to 25) are characterized by SJV cluster 3, low ozone in the SFB and high ozone in the SV. The worst-case scenario in summer 2000 **days 211 to 215** (July 29 to August 2) is largely aligned with high SJV ozone (cluster 6) levels, high SFB ozone, and moderate SV ozone. These characteristic episodes are largely coincident with the Intensive Operation Periods (IOPs, see Fujita et al. 2001), except the June episode. Modeling these IOP days may reasonably represent most of the temperature and flow conditions.

Comparing to 30-year historical climatology, Fujita et al. (2001) showed that inland temperatures in the CCOS domain during summer 2000 were statistically cooler than normal. Fresno average daily temperature maxima (306.0 K) was more than three standard deviations below the climatological value of 307.9 K. Accordingly, our analysis found the most stagnant conditions (cluster 4) occurred least frequently, although it was the most frequent cluster in the historical data set. Air quality modeling efforts for summer 2000 focused on the high-ozone episode (July 29 to August 2), which has the highest temperature levels in the CCOS domain (about 2–3 K higher than SJV clusters 3 to 5). Modeling this episode alone may represent pollution production rates under typical high-temperature conditions. Flow patterns in this episode are not representative of those in other clusters, especially SJV cluster 4 (historical cluster 1), which has the most stagnant conditions. Simulation results from this episode would tend to over-estimate inter-basin contributions of pollutants from the SFB for SJV cluster 4, but under-estimate it for SJV cluster 5. Ideally, ozone sensitivity studies should be conducted for all the episode types, to determine the variability with respect to inter-basin contributions, and whether it is significantly different across flow regimes associated with high ozone clusters. Finally, since summer 2000 was cooler than usual, the control strategies determined from this modeling period should be examined with second-order sensitivity analysis by examining the influence of temperature changes on ozone responses to emission reductions.

CHAPTER 6:
Conclusions and Recommendations

Improved analysis and prediction of air quality impacts from emissions sources such as power plants is required for planners and regulators to more accurately determine the true impact of various sources on air quality and evaluate alternative control strategies such as short-term curtailment of emissions. This research has considered some of the limitations of modeling just a few high-ozone episodes as the basis for these types of planning and evaluations. The results of this study indicate that high ozone occurs during multiple types of summer meteorological flow patterns, underscoring the need for a more inclusive approach than episodic modeling. Some of the specific findings and recommendations are discussed below.

6.1 Summary of Major Findings

In this research, we applied the Community Multiscale Air Quality Model (CMAQ) to simulate air quality in Central California for an entire summer season, and characterized ozone patterns and their associated meteorological conditions with cluster analysis.

An important result of this study is a suite of benchmarked model inputs that describe emissions and meteorological variations in both time and space for Central California for an entire summer season. A variety of evaluation methods and diagnostic tools have been applied to refine model inputs and to evaluate model performance. Initially this was done for a 15-day period, then the modeling period was extended to the entire summer of 2000. Lateral boundary conditions, minimum vertical diffusivity, and ozone dry deposition rate over the ocean are refined according to model diagnostics, observations, and the published literature. Gridded meteorological and emission inputs are developed to reflect variability occurring on diurnal, weekly, and seasonal time scales. Driven by these inputs, the model has stable performance for the entire modeling period without further adjustment of input parameters. The model is able to reproduce ozone and its precursor species concentrations observed in the San Joaquin Valley; however, model performance in other parts of the domain, especially the San Francisco Bay area, still needs improvement.

A high-order direct decoupled method (HDDM) for sensitivity analysis is demonstrated to provide an effective tool to delineate ozone responses to precursor reductions in time and space, and to determine the relative importance of emission contributions to ozone concentrations from local versus upwind areas. Influences of uncertainties in model input parameters on ozone responses to emission changes are determined using second-order sensitivity analyses. For the conditions analyzed here, nitrogen oxide (NOx) emission reductions were overall more effective than reductions of volatile organic compounds (VOCs) for attaining the 8-h ozone standard in the San Joaquin Valley. In contrast, VOC reductions would have worked better for attaining the prior 1-h ozone standard. Inter-basin source contributions of NOx emissions are limited to the northern part of the SJV, while anthropogenic VOC (AVOC) emissions, especially those emitted at night, influence ozone formation in the SJV further downwind. Among model input parameters studied here, uncertainties in emissions of NOx and AVOC, and the rate coefficient

of the $OH+NO_2$ termination reaction, are found to have the greatest effects on first-order ozone responses to changes in NOx emissions. Uncertainties in biogenic VOC emissions only have a modest effect because they are generally not collocated with anthropogenic sources in the SJV.

Using cluster analysis and principal component analysis, we have determined distinctive meteorological regimes that are associated with different ozone spatial patterns in Central California. Meteorological effects are shown to explain the observed ozone spatial patterns, their dynamic relationship with each other, and their relationship to those in upwind areas. In general, average ozone levels increase with temperature, while their spatial distributions depend on flow regimes, especially the strength of sea breezes and upslope flows. The high-ozone episode in late July/early August of 2000 is shown to have a different flow regime that is distinct from other days in summer 2000, when significant ozone exceedances occur under more stagnant conditions. Ozone spatial patterns and the associated weather regimes captured in these simulations are reflected in those derived from a historical observational data set, which provides an indication of representativeness for the modeling period, and indicates that the model reproduces observed ozone patterns for the right reasons.

6.2 Recommendations for Future Research

Many of the protocols used in air quality modeling were dictated by a lack of computer resources that have been largely removed. These are separating the meteorology calculations from the chemical transport calculations, examining episodes rather than longer periods, modeling at lower resolution, and limiting spatial domain. These practices need to change and keep pace with the times.

This research shows the advantages of modeling a full summer ozone season at a high spatial resolution. The high-ozone episode discussed in Chapter 2 was found to have different temperature and flow patterns compared to days assigned to the other high-ozone clusters in summer 2000. Variability in ozone responses to emission changes and inter-basin contributions should be quantified by extending the HDDM sensitivity analysis to days associated with the other types of high ozone, to achieve a more comprehensive and effective basis for control strategy design. Studies can be performed with updated chemical mechanisms (Shearer et al. 2012). Since the summer of 2000 was cooler than usual, ozone production rates in this period may not be representative of those for more typical (warmer) summers. Ozone sensitivities derived from this modeling period should be examined with second-order sensitivity analysis to determine the influence of temperature changes on ozone responses to emission reductions.

Weather patterns associated with different ozone clusters were examined only with surface meteorological variables in this research, which assumes that the prevailing meteorological regimes have significantly affected the ground-level data, from which the weather patterns can be inferred. A more complete description of mesoscale and synoptic scale regimes can be determined by examining upper-level atmospheric conditions. As an example, the evolution and localization of certain high-pressure systems may help to explain the variability within an ozone cluster and its dynamic relationship to the other ozone clusters to which it transitions.

Effectiveness of emission-control strategies reflected in changes of ambient concentrations can be easily confounded by changes in meteorology. A more objective evaluation can be conducted by examining ozone changes for different emission scenarios under similar meteorological conditions. The meteorological features identified in this research that determine ozone levels and spatial extents could be used to identify meteorological similarity or lack of similarity between two time periods.

The statistical significance of speculations made with respect to ozone clusters such as the occurrence frequency, transition probability, and relationship to upwind ozone patterns can be further improved by simulating more summer seasons. Multiple years are not routinely simulated, but it could be done if there were enough observations to help establish and evaluate model performance. Findings from research where multiple years were simulated and ozone spatial patterns and important meteorological features were identified could be used to guide future observational studies. Changes in wind speeds and directions along the coastal passes and the western slope of the Sierra Nevada are found to produce the largest variabilities in regional wind fields. Thus meteorological monitoring sites in these locations should be given more weight in prioritizing flow regimes that influence ozone. Although hourly wind measurements alone can decide important regimes that affect air quality in the Bay area (Beaver and Palazoglu 2006b), this research has shown that it is not the case for the San Joaquin Valley. Different ozone levels in the SJV occur under similar flow regimes. Both temperature distributions and wind measurements should be considered when identifying important weather regimes that lead to high ozone concentrations. Existing ozone monitoring sites can reasonably capture various ozone spatial patterns in the Bay area and the Sacramento Valley, but cannot distinguish ozone patterns in the SJV under different stagnant conditions. Days with these different SJV ozone and weather patterns could be separated if additional measurements were conducted along the western side of the SJV, or by including the SFB ozone observations into the clustering procedure.

By clustering modeled data, this research has depicted ozone spatial patterns within the San Joaquin Valley under different meteorological conditions. Ozone levels on the western side of the SJV can change significantly, depending on the degree of atmospheric stagnation that is characterized by the strength of sea breeze and ventilation flows to the Sierra Nevada. This spatial feature has not been captured by the existing ozone monitoring stations, which are located in the more populated middle and eastern side of the SJV. The western part of the SJV is dominated by agriculture, where crop productivity is adversely affected by ambient ozone levels (e.g., Morgan et al. 2003). Ozone levels in this area are affected both by motor vehicles traveling on Highway 5 and by pollutants transported from the Bay area and other coastal regions. Additional measurements of pollutants along the western part of the SJV could be used to enhance understanding of health effects and crop damage by ozone and its precursors, and to assess changes in ozone concentrations brought about by reducing emissions from local and upwind sources.

Model performance evaluation has shown that ozone predictions in the Bay area and Sacramento Valley agree less satisfactory with observations in episodes that are subject to the largest uncertainties in meteorological fields (wind directions in the Sacramento Valley and

timing and strength of onshore flow in the Bay area) within the boundary layer. Currently, wind profilers deployed in Central California have more spatial coverage in the SJV than other air basins. More meteorological measurements are needed to improve the understanding and characterization of wind flow and other boundary layer dynamics in the SV and SFB, which can subsequently help the ozone modeling.

Emission inputs have been considered as one of the most important, yet most uncertain, inputs to air quality models (Russell and Dennis 2000). Emission inventories need careful evaluation due to their considerable complexity. Greater attention should be devoted to quantifying uncertainties. Anthropogenic emissions can be affected by meteorological factors; for example, evaporative VOC emissions increase with ambient temperature (Rubin et al. 2006). In the current research, the same weekly cycle was used to simulate ozone for the entire summer. The anthropogenic emission inputs could be improved by imposing day-specific variability that is driven by meteorology. Such variability could be important for simulating ozone for an entire summer, during which significant temperature changes occur. Human activities during the summer could also be different from month to month and from day to day. Factors that affect anthropogenic emissions include periods of intensive harvest, vacations, and other changes in economic activities. The diurnal emission profiles used to allocate total daily emissions to different hours of the day should be investigated to address the unexplained positive bias in ozone predictions at night, after titration effects have been accounted for. Improvements in emission inventories could be achieved through a "bottom-up" approach, where more accurate source activities are estimated using ambient observations. Possible biases in emission inventories could also be identified through a "top-down" approach, where precursor sensitivities to emission perturbations are used to relate biases in concentration predictions to those in emission inventories (e.g., Napelenok et al. 2008).

The model evaluation conducted in this research for Central California indicates that model performance in the San Francisco Bay area and coastal areas is consistently less satisfactory when compared to inland locations. In particular, the model overpredicts ozone concentrations in coastal areas even when transport errors are minimized by wind nudging. As a result of proximity to the Pacific ocean, coastal sites are influenced by a stable marine boundary layer and advective clouds/fog, which lead to different vertical mixing and actinic fluxes relative to those inland. Investigations of CMAQ model parameterization schemes that are related to vertical mixing and actinic flux should be conducted to identify areas for further model improvement.

GLOSSARY

Term	Definition
ABK	Arbuckle
ABL	Atmospheric Boundary Layer
ADE	Atmospheric Diffusion Equation
AGL	Above Ground Level
AGO	Angiola
APCD	Air Pollution Control District
AQMD	Air Quality Management District
ARB	California Air Resources Board
AVOC	Anthropogenic Volatile Organic Compounds
BKF	Bakersfield
BVOC	Biogenic Volatile Organic Compounds
CCHO	Acetaldehyde
CCO	Chico
CCOS	Central California Ozone Study
CH4	Methane
CMAQ	Community Multiscale Air Quality Model
CO	Carbon Monoxide
CTM	Chemical Transport Model
DDM	Directed Decoupled Method
EBI	Euler Backward Iterative
EES	Emissions Estimation System
EKMA	Empirical Kinetic Modeling Approach
EOF	Empirical Orthogonal Function
EPA	U.S. Environmental Protection Agency
Eta ABL	Eta Atmospheric Boundary Level

FAT	Fresno
HCHO	Formaldehyde
HDDM	High-Order Direct Decoupled Method
IOPs	Intensive Operating Periods
JPL	NASA Jet Propulsion Laboratory
LBA	Los Banos
LEM	Lemoore
LHS	Lost Hills
LVR	Livermore
MC	Mountain Counties
MCIP	Meteorology to Chemistry Interface Processor
MM5	Mesoscale Model
N_2O	Nitrous Oxide
NCC	North Central Coast
NCEP	National Center for Environmental Prediction
NH_3	Ammonia
NMHC	Non-Methane Hydro Carbon
NMOC	Non-Methane Organic Carbon
NOAA	National Oceanic Atmospheric Administration
NOx	Nitrogen Oxides
OPE	Ozone Production Efficiency
PAN	Peroxyacetyl Nitrate
PC	Principal Component
PCA	Principal Component Analysis
PDT	Pacific Daylight Time
PIER	Public Interest Energy Research
$PM_{2.5}$	Particulate Matter ≤ 2.5 micrometers,

PM_{10}	Particulate matter ≤ 10 micrometers
ppb	Parts per billion
ppbC	Parts per billion Carbon
PSG	Pleasant Grove
RMSD	Root Mean Square Difference
ROG	Reactive Organic Gases
RRTM	Rapid Radiation Transfer Model
SAC	Sacramento
SAQM	SARMAP Air Quality Model
SARMAP	SJVAQS/AUSPEX Regional Modeling Adaptation Project
SCC	South Central Coast
SFB	San Francisco Bay
SIP	State Implementation Plans
SJV	San Joaquin Valley
SMVGEAR	Sparse Matrix Vectorized Gear
SO_2	Sulfur Dioxide
SV	Sacramento Valley
TCY	Tracy
TRA	Travis
UTC	Coordinated Universal Time
VOCs	Volatile Organic Compounds
WFD	Waterford
WSS	Within-group Sum of Squares

REFERENCES

Anderberg, M. R. (1973). *Cluster Analysis for Applications. Monographs and Textbooks on Probability and Mathematical Statistics.* New York: Academic Press Inc. 359 pp.

Baertsch-Ritter, N., Keller, J. Dommen, J. and Prevot, A. S. H. (2004). "Effects of various meteorological conditions and spatial emission resolutions on the ozone concentration and ROG/NOx limitation in the Milan area (I)." *Atmospheric Chemistry and Physics* 4: 423–438.

Bao, J. W., Michelson, S. A., Persson, O., Bianco, L., Djalalova, I., White, D. E. and Wilczak, J. M. (2006). *NOAA Contributions to the Central California Ozone Study and Ongoing Meteorological Monitoring.* Draft final report, prepared for the San Joaquin Valleywide Air Pollution Study Agency, California Air Resources Board, by NOAA/Earth Systems Research Laboratory, Boulder, Colorado. 25 November.

Beaver, S. and Palazoglu, A. (2006a). "A cluster aggregation scheme for ozone episode selection in the San Francisco, CA Bay Area." *Atmospheric Environment* 40(4): 713–725.

Beaver, S. and Palazoglu, A. (2006b). "Cluster analysis of hourly wind measurements to reveal synoptic regimes affecting air quality." *Journal of Applied Meteorology and Climatology* 45(12): 1710–1726.

Beaver, S. Palazoglu, A. et al. (2008). Cluster sequencing to analyze synoptic transitions affecting regional ozone. *Journal of Applied Meteorology and Climatology* 47(3): 901–916.

Beaver, S. and Palazoglu, A. (2009). Influence of synoptic and mesoscale meteorology on ozone pollution potential for San Joaquin Valley of California. *Atmospheric Environment* 43(10): 1779–1788.

Bell, M. L. and Dominici, F. (2008). "Effect Modification by Community Characteristics on the Short-term Effects of Ozone Exposure and Mortality in 98 US Communities." *Am. J. Epidemiol.* 167(8): 986–997.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M. and Dominici, F. (2004). "Ozone and Short-term Mortality in 95 US Urban Communities, 1987–2000." *JAMA* 292(19): 2372–2378.

Biswas, J., Hogrefe, C., Rao, S. T., Hao, W. and Sistla, G. (2001). "Evaluating the performance of regional-scale photochemical modeling systems. Part III--Precursor predictions." *Atmospheric Environment* 35(35): 6129

Blanchard, C. L. and Stoeckenius, T. (2001). "Ozone response to precursor controls: Comparison of data analysis methods with the predictions of photochemical air quality simulation models." *Atmospheric Environment* 35(7): 1203–1215.

Brown, N. J. and Revzan, K. L. (2005). "Comparative sensitivity analysis of transport properties and reaction rate coefficients." *International Journal of Chemical Kinetics* 37(9): 538–553.

Byun, D. W. and Schere, K. L. (2006). "Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system." *Applied Mechanics Reviews* 59(2): 51–77.

Carter, W. P. L. (2000). *Implementation of the SARPRC-99 Chemical Mechanism into the Model-3 Framework.* University of California, Riverside, California, Washington, D.C.

Coats, C. J. (1995). High Performance Algorithms in the Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System, Microelectrics Center of North Carolina, Environmental Systems Division, Research Triangle Park, North Carolina.

Cohan, D. S., Hakami, A., Hu, Y. T. and Russell, A. G. (2005). "Nonlinear response of ozone to emissions: Source apportionment and sensitivity analysis." *Environmental Science & Technology* 39(17): 6739–6748.

Cohn, R. D., Eder, B. K., Leduc, S. K. and Dennis, R. L. (2001). "Development of an Aggregation and Episode Selection Scheme to Support the Models-3 Community Multiscale Air Quality Model." *Journal of Applied Meteorology* 40(2): 210–228.

Couach, O., Kirchner, F., Jimenez, R., Balin, I., Perego, S. and van den Bergh, H. (2004). "A development of ozone abatement strategies for the Grenoble area using modeling and indicators." *Atmospheric Environment* 38(10): 1425–1436.

Cressie, N. A. (1993). *Statistics for Spatial Data.* Revised edition. New York: John Wiley and Sons.

Crutcher, H. L., Rhodes, R. C., Graves, M. E., Fairbairn, B. and Nelson, A. C. (1986). "Application of cluster analysis to aerometric data." *Journal of Air Pollution Control Association* 36: 1116–1122.

Dabdub, D., DeHaan, L. L. and Seinfeld, J. H. (1999). "Analysis of ozone in the San Joaquin Valley of California." *Atmospheric Environment* 33(16): 2501–2514.

Darby, L. S. (2005). "Cluster analysis of surface winds in Houston, Texas, and the impact of wind patterns on ozone." *Journal of Applied Meteorology* 44(12): 1788–1806.

Dawson, J. P., Adams, P. J. and Pandis, S. N. (2007). "Sensitivity of ozone to summertime climate in the eastern USA: A modeling case study." *Atmospheric Environment* 41(7): 1494.

de Foy, B., Fast, J. D., Paech, S. J., Phillips, D., Walters, J. T., Coulter, R. L., Martin, T. J., Pekour, M. S., Shaw, W. J., Kastendeuch, P. P., Marley, N. A., Retama, A. and Molina, L. T. (2008). "Basin-scale wind transport during the MILAGRO field campaign and comparison to climatology using cluster analysis." *Atmospheric Chemistry and Physics* 8(5): 1209–1224.

Dommenget, D. and Latif, M. (2002). "A cautionary note on the interpretation of EOFs." *Journal of Climate* 15(2): 216–225.

Dunker, A. M. (1981). "Efficient Calculation of Sensitivity Coefficients for Complex Atmospheric Models." *Atmospheric Environment* 15(7): 1155–1161.

Dunker, A. M., Yarwood, G., Ortmann, J. P. and Wilson, G. M. (2002). "The decoupled direct method for sensitivity analysis in a three-dimensional air quality model - Implementation, accuracy, and efficiency." *Environmental Science & Technology* 36(13): 2965–2976.

Eder, B. K., Davis, J. M. and Bloomfield, P. (1994). An Automated Classification Scheme Designed to Better Elucidate the Dependence of Ozone on Meteorology. *Journal of Applied Meteorology* Vol. 33(No. 10): 1182–1199.

Eder, B., Kang, D., Mathur, R., Yu, S. and Schere, K. (2006). "An operational evaluation of the Eta-CMAQ air quality forecast model." *Atmospheric Environment* 40(26): 4894.

Eder, B., Kang, D., Mathur, R., Pleim, J, Yu, S., Otte, T, and Pouliot, G. (2009). "A performance evaluation of the National Air Quality Forecast Capability for the summer of 2007." *Atmospheric Environment* 43(14): 2312.

Everitt, B. S. (1993). *Cluster Analysis*. 3rd ed. London: Heinemann Education.

Everitt, B. S., Landau, S. and Leese, M. (2001). *Cluster Analysis.* 4th ed. Edward Arnold Publishers. 229 pp.

FEJF (Fire Emissions Joint Forum) (1996). *Fire Emission Inventory for the WRAP Region-Methodology and Emission Estimates.* Final Report by Air Sciences Inc. for the Western Regional Air Partnership. 1996.

Fernau, M. E. and Samson, P. J. (1990). "Use of Cluster-Analysis to Define Periods of Similar Meteorology and Precipitation Chemistry in Eastern North-America. 1. Transport Patterns." *Journal of Applied Meteorology* 29(8): 735–750.

Fujita, E., Keislar, R., Stockwell, W., Moosuller, H., DuBois, D., Koracin, D. and Zielinska, B. (1999). Central California Ozone Study-Volume I, Field Study Plan. Division of Atmospheric Science Desert Research Institute, 2215 Raggio Parkway, Reno, Nevada.

Fujita, E., D. Campbell, R. Keisler, J. Brown, S. Tanrikulu, and A. J. Ranzieri. (2001). *Central California Ozone Study (CCOS) - Final report, Volume III: Summary of field operations.* Technical report. California Air Resources Board.

Galvez, O. (2007). "Synoptic-scale transport of ozone into Southern Ontario." *Atmospheric Environment* 41: 8579–8595.

Gego, E., Gilliland, A., Godowitch, J., Rao, S.T., Porter, P. S. and Hogrefe, C. (2008). "Modeling Analyses of the Effects of Changes in Nitrogen Oxides Emissions from the Electric Power Sector on Ozone Levels in the Eastern United States." *Journal of the Air & Waste Management Association* 58(4): 580–588.

Goldstein, A. H., D. B. Millet, M. McKay, L. Jaegle, L. Horowitz, O. Cooper, R. Hudman, D. J. Jacob, S. Oltmans, and A. Clarke (2004). "Impact of Asian emissions on observations at Trinidad Head, California, during ITCT 2K2." *J. Geophys. Res.* 109: doi:10.1029/2003JD004406.

Grell, G. A. (1993). "Prognostic evaluation of assumptions used by cumulus parameterizations." *Mon. Wea. Rev.* 121: 764–787.

Grell, G. A., Dudhia, J. and Stauffer, D. R. (1994). A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR Technical Note, NCAR/TN-398+STR, 122 pp., NCAR, Boulder, Colorado.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J. and Zimmerman, P. (1995). "A Global-Model of Natural Volatile Organic-Compound Emissions." *Journal of Geophysical Research-Atmospheres* 100(D5): 8873–8892.

Hakami, A., Odman, M. T. and Russell, A. G. (2003). "High-order, direct sensitivity analysis of multidimensional air quality models." *Environmental Science & Technology* 37(11): 2442–2452.

Hakami, A., Odman, M. T. and Russell, A. G. (2004). "Nonlinearity in atmospheric response: A direct sensitivity analysis approach." *Journal of Geophysical Research-Atmospheres* 109(D15): D15303, doi:10.1029/2003JD004502.

Harley, R. A., Marr, L. C., Lehner, J. K. and Giddings, S. N. (2005). "Changes in motor vehicle emissions on diurnal to decadal time scales and effects on atmospheric composition." *Environmental Science & Technology* 39(14): 5356–5362.

Harley, R. A., Brown, N. J., Tonse, S. R., and Jin, L. (2006). *A seasonal perspective on regional air quality in Central California.* Department of Civil and Environmental Engineering. UC Berkeley and Atmospheric Science Department, Lawrence Berkeley National Laboratory, Berkeley, California. Phase I final report for San Joaquin Valley wide Air Pollution Study Agency and the California Air Resources Board.

Hogrefe, C., Rao, S. T., Kasibhatla, P., Hao, W., Sistla, G., Mathur, R. and McHenry, J. (2001). "Evaluating the performance of regional-scale photochemical modeling systems: Part II--ozone predictions." *Atmospheric Environment* 35(24): 4175.

Hogrefe, C., Biswas, J., Lynn, B., Civerolo, K., Ku, J. Y., Rosenthal, J., Rosenzweig, C., Goldberg, R. and Kinney, P. L. (2004a). "Simulating regional-scale ozone climatology over the eastern United States: Model evaluation results." *Atmospheric Environment* 38(17): 2627.

Hogrefe, C., Porter, P. S., Gego, E., Gilliland, A., Gilliam, R., Swall, J., Irwin, J. and Rao, S. T. (2006). "Temporal features in observed and simulated meteorology and air quality over the Eastern United States." *Atmospheric Environment* 40(26): 5041.

Hotelling, H. (1935). "The most predictable criterion." *J. Ed. Psych.* 26: 139–142.

Janjić, Z. I. (2002). Nonsingular Implementation of the Mellor–Yamada Level 2.5 Scheme in the NCEP Meso model. NCEP Office Note, No. 437. 61 pp.

Jin, L., Tonse, S., Cohan, D. S., Mao, X., Harley, R. A. and Brown, N. J. (2008). "Sensitivity Analysis of Ozone Formation and Transport for a Central California Air Pollution Episode." *Environ. Sci. Technol.* 42(10): 3683–3689.

Jin, L., N. J. Brown, R. A. Harley, J-W. Bao, S. Michelson, and J. M. Wilzak. (2010). Seasonal Versus Episodic Performance Evaluation for an Eulerian Photochemical Air Quality Model. *Journal of Geophysical Research Atmospheres* 115: D09302,doi:10.1029/2009JDO12680.

Jin, L., R. A. Harley, N. J. Brown (2011). Ozone pollution regimes modeled for a summer season in California's San Joaquin Valley: A cluster analysis. *Atmospheric Environment* 45(27): 4707-4718.

Kaduwela, A. (2006). Affiliation at California Air Resources Board, personal communication.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley & Sons, Inc.

Kaufmann, P. and Weber, R. O. (1996). "Classification of Mesoscale Wind Fields in the MISTRAL Field Experiment." *Journal of Applied Meteorology* 35(11): 1963–1979.

Liao, D., Peuquet, D., Duan, Y., Whitsel, E., Dou, J., Smith, R., Lin, H.-M., Chen, J.-C. and Heiss, G. (2006). GIS approaches for the estimation of residential level ambient PM concentrations. *Environ. Health Perspect.* **114**(9): 1374-1380.

Lin, Y. L. and Jao, I. C. (1995). A Numerical Study of Flow Circulations in the Central Valley of California and Formation Mechanisms of the Fresno Eddy. *Monthly Weather Review* 123(11): 3227–3239.

Lippmann, M. (1989). "Health effects of ozone. A critical review." *JAPCA* 39(5): 672–695.

Lippmann, M. (1993). "Health effects of tropospheric ozone: Review of recent research findings and their implications to ambient air quality standards." *J Expo Anal Environ Epidemiol* 3(1): 103–29.

Loomis, C. (2003). *Development of Stack Parameters and Vertical Distributions for Modeling Large Wildfires in the CCOS Domain.* Alpine Geophysics Inc.

Lorenz, E. N. (1956). *Empirical orthogonal functions and statistical weather prediction.* Technical report. Statistical Forecast Project Report 1, Dept. of Meteor. MIT, 49 pp.

Ludwig, F. L., Jiang, J.-Y. and Chen, J. (1995). "Classification of ozone and weather patterns associated with high ozone concentrations in the San Francisco and Monterey Bay areas." *Atmospheric Environment* 29(21): 2915.

Martien, P. T. and Harley, R. A. (2006). "Adjoint sensitivity analysis for a three-dimensional photochemical model: Application to Southern California." *Environmental Science & Technology* 40(13): 4200–4210.

Matern, B. (1960). Spatial Variation—Stochastic Models and Their Application to some Problems in Forest Surveys and other Sampling Investigations. Stockholm: Medd. Statens Skogsforskningsinstitut. 49 no. 5.

Mebust, M. R., Eder, B. K., Binkowski, F. S. and Roselle, S. J. (2003). "Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 2. Model evaluation." *J. Geophys. Res.* 108 doi:10.1029/2001JD001410.

Mena-Carrasco, M., Tang, Y., Carmichael, G. R., Chai, T., Thongbongchoo, N., Campbell, J. E., Kulkarni, S., Horowitz, L., Vukovich, J., Avery, M., Brune, W., Dibb, J. E., Emmons, L., Flocke, F., Sachse, G. W., Tan, D., Shetter, R., Talbot, R. W., Streets, D. G., Frost, G. and Blake, D. (2007). "Improving regional ozone modeling through systematic evaluation of errors using the aircraft observations during the International Consortium for Atmospheric Research on Transport and Transformation." *J. Geophys. Res.* 112: D12S19, doi:10.1029/2006JD007762.

Michelson, S. A., I. V. Djalalova, and J.-W. Bao. (2009). Evaluation of the summertime low-level winds simulated by MM5 in the Central Valley of California. Submitted to *J. of Appl. Meteor. Clim.*

Morgan, P. B., Ainsworth, E. A. and Long, S. P. (2003). "How does elevated ozone impact soybean? A meta-analysis of photosynthesis, growth and yield." *Plant, Cell & Environment* 26(8): 1317–1328.

Napelenok, S. L., Cohan, D. S., Hu, Y. and Russell, A. G. (2006). "Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3D/PM)." *Atmospheric Environment* 40(32): 6112–6121.

Napelenok, S. L., Cohan, D. S. et al. (2008). "Extension and evaluation of sensitivity analysis capabilities in a photochemical model." *Environ. Model. Softw* 23(8): 994–999.

Napelenok, S. L., Pinder, R. W., Gilliland, A. B. and Martin, R. V. (2008). "A method for evaluating spatially-resolved NOx emissions using Kalman filter inversion, direct sensitivities, and space-based NO_2 observations." *Atmos. Chem. Phys. Discuss.* 8(2): 6469.

Newchurch, M. J., Ayoub, M. A., Oltmans, S., Johnson, B. and Schmidlin, F. J. (2003). "Vertical distribution of ozone at four sites in the United States." *Journal of Geophysical Research-Atmospheres* 108(D1): D18303, doi:10.1029/2005JD006935.

Nowak, J. B., D. D. Parrish, J. A. Neuman, J. S. Holloway, O. R. Cooper, T. B. Ryerson, J. D. K. Nicks, F. Flocke, J. M. Roberts, E. Atlas, J. A. d. Gouw, S. Donnelly, E. Dunlea, G. H, L. G. Huey, S. Schauffler, D. T. Sueper, D. J. Tanner, C. Warneke, F. C. Fehsenfeld (2004). "Gas-phase chemical characteristics of Asian emission plumes observed during ITCT 2K2 over the eastern North Pacific Ocean." *J. Geophys. Res.* 109: D23S19, doi:10.1029/2003JD004488.

Oliver, M. A. and Webster, R. (1990). "Kriging: A method of interpolation for geographical information systems." *Int. J. Geogr. Inf. Syst.* 4(3): 313–332.

Pearson, K. (1902). "On lines and planes of closest fit to systems of points in space." *Phil. Mag.* 2: 559–572.

Pun, B. K., Louis, J. F., Pai, P., Seigneur, C., Altshuler, S. and Franco, G. (2000). "Ozone formation in California's San Joaquin Valley: A critical assessment of modeling and data needs." *Journal of the Air & Waste Management Association* 50(6): 961–971.

Reynolds, S. D., Blanchard, C. L. and Ziman, S. D. (2003). "Understanding the effectiveness of precursor reductions in lowering 8-hr ozone concentrations." *Journal of the Air & Waste Management Association* 53(2): 195–205.

Rubin, J., Kean, A. J., Harley, R. A., Millet, D. B. and Goldstein, A. H. (2006). "Temperature dependence of volatile organic compound evaporative emissions from motor vehicles." *Journal of Geophysical Research-Atmospheres* 111: D03305, doi:10.1029/2005JD006458.

Russell, A. and Dennis, R. (2000). "NARSTO critical review of photochemical models and modeling." *Atmospheric Environment* 34(12-14): 2283.

Sander, S. P., Friedl, R. R., Ravishankara, A. R., Golden, D. M., Kolb, C. E., Kurylo, M. J., Molina, M. J., Moortgat, G. K., Keller-Rudek, H., Finlayson-Pitts, B. J., Wine, P. H., Huie, R. E. and Orkin, V. L. (2006). Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies. Jet Propulsion Laboratory, Pasadena, California.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K. and Chatterton, T. (2003). "A rigorous inter-comparison of ground-level ozone predictions." *Atmospheric Environment* 37(23): 3237–3253.

Schwarzhoff, P. J. and Reid, P. D. (2000). "Classification of meteorological patterns associated with the ozone categories in Kelowna, British Columbia." *Journal of Applied Meteorology* 39(3): 463–470.

Scott, K. I. and Benjamin, M. T. (2003). "Development of a biogenic volatile organic compounds emission inventory for the SCOS97-NARSTO domain." *Atmospheric Environment* 37(Supplement 2): 39–49.

Shearer, S. M., Harley, R. A., Jin, L. and Brown, N. J. (2012). "Comparison of SAPRC99 and SAPRC07 mechanisms in photochemical modeling for central California." *Atmospheric Environment* 46(0): 205–216.

Sillman S., J. A. L. and Wofsy, S. C. (1995). "The sensitivity of ozone to nitrogen oxides and hydrocarbons in urban locations." *J. Geophys. Res.* 100: 14175–14188.

Singh, H., Y. Chen, A. Staudt, D. Jacob, D. Blake, B. Heikes and J. Snow (2001). "Evidence from the Pacific troposphere for large global sources of oxygenated organic compounds." *Nature* 410: 1078–1081.

Stein, M. L. (1999). *Interpolation of Spatial Data. Some Theory for Kriging.* New York: Springer. 247pp.

Steiner, A. L., S. Tonse, R. C. Cohen, A. H. Goldstein, R. A. Harley (2006). "Influence of future climate and emissions on regional air quality in California." *J. Geophys. Res.* 111, D18303, doi:10.1029/2005JD006935.

Stevenson, D. S., Johnson, C. E., Collins, W. J., Derwent, R. G. and Edwards, J. M. (2000). "Future estimates of tropospheric ozone radiative forcing and methane turnover - the impact of climate change." *Geophysical Research Letters* 27(14): 2073–2076.

Swall, J. L. and Davis, J. M. (2006). "A Bayesian statistical approach for the evaluation of CMAQ." *Atmospheric Environment* 40(26): 4883.

Swall, J. L. and Foley, K. M. (2009). "The impact of spatial correlation and incommensurability on model evaluation." *Atmospheric Environment* 43(6): 1204.

Tao, Z., Williams, A., Huang, H.-C., Caughey, M. and Liang, X.-Z. (2007). "Sensitivity of U.S. surface ozone to future emissions and climate changes." *Geophys. Res. Lett.* 34.

Taylor, K. E. (2001). "Summarizing multiple aspects of model performance in a single diagram." *J. Geophys. Res.* 106(D7): 7183–7192.

Tonse, S., Brown, N.J., Jin, L., Harley, R.A. (2008). "A Process-Analysis Based Study of the Ozone Weekend Effect." *Atmospheric Environment* 42: 7728–7736.

Thompson, M. L., Reynolds, J., Cox, L. H., Guttorp, P. and Sampson, P. D. (2001). "A review of statistical methods for the meteorological adjustment of tropospheric ozone." *Atmospheric Environment* 35(3): 617–630.

Tong, D. Q. and Mauzerall, D. L. (2006). "Spatial variability of summertime tropospheric ozone over the continental United States: Implications of an evaluation of the CMAQ model." *Atmospheric Environment* 40(17): 3041.

Tonse, S., Brown, N.J., Jin, L., Harley, R.A. (2008). "A Process-Analysis Based Study of the Ozone Weekend Effect." *Atmospheric Environment* 42: 7728–7736.

Trainer, M., Parrish, D. D., Buhr, M. P., Norton, R. B., Fehsenfeld, F. C., Anlauf, K. G., Bottenheim, J. W., Tang, Y. Z., Wiebe, H. A., Roberts, J. M., Tanner, R. L., Newman, L., Bowersox, V. C., Meagher, J. F., Olszyna, K. J., Rodgers, M. O., Wang, T., Berresheim, H., Demerjian, K. L. and Roychowdhury, U. K. (1993). "Correlation of Ozone within NOy in Photochemically Aged Air." *J. Geophys. Res.* 98(D2): 2917–2925.

Trainer, M., Parrish, D. D., Goldan, P. D., Roberts, J. and Fehsenfeld, F. C. (2000). "Review of observation-based analysis of the regional factors influencing ozone concentrations." *Atmospheric Environment* 34(12-14): 2045–2061.

U.S. EPA (1991). Guideline for Regulatory Application of the Urban Airshed Model. EPA450/4-91-013. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

U.S. EPA (1994). Clean Air Act Ozone Design Value Study: Final Report, A Report to Congress EPA-454/R-94-035. Office of Air Quality Planning and Standards. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. December.

U.S. EPA (2008). National Ambient Air Quality Standards for Ozone. Federal Register, Vol. 73, No. 60. Rules and Regulations. Thursday, March 27.

Vuilleumier, L., Harley, R. A. and Brown, N. J. (1997). "First- and second-order sensitivity analysis of a photochemically reactive system (a Green's function approach)." *Environmental Science & Technology* 31(4): 1206–1217.

Wackernagel, H., Lajaunie, C., Blond, N., Roth, C. and Vautard, R. (2004). "Geostatistical risk mapping with chemical transport model output and ozone station data." *Ecological Modelling* 179(2): 177.

Winner, D. A. and Cass, G. R. (2000). "Effect of Emissions Control on the Long-Term Frequency Distribution of Regional Ozone Concentrations." *Environ. Sci. Technol.* 34(12): 2612–2617.

Yang, Y. J., Wilkinson, J. G. and Russell, A. G. (1997). "Fast, Direct Sensitivity Analysis of Multidimensional Photochemical Models." *Environ. Sci. Technol.* 31(10): 2859–2868.

Yarnal, B. (1993). *Synoptic Climatology in Environmental Analysis.* 1st ed. Belhaven Press. 195 pp.

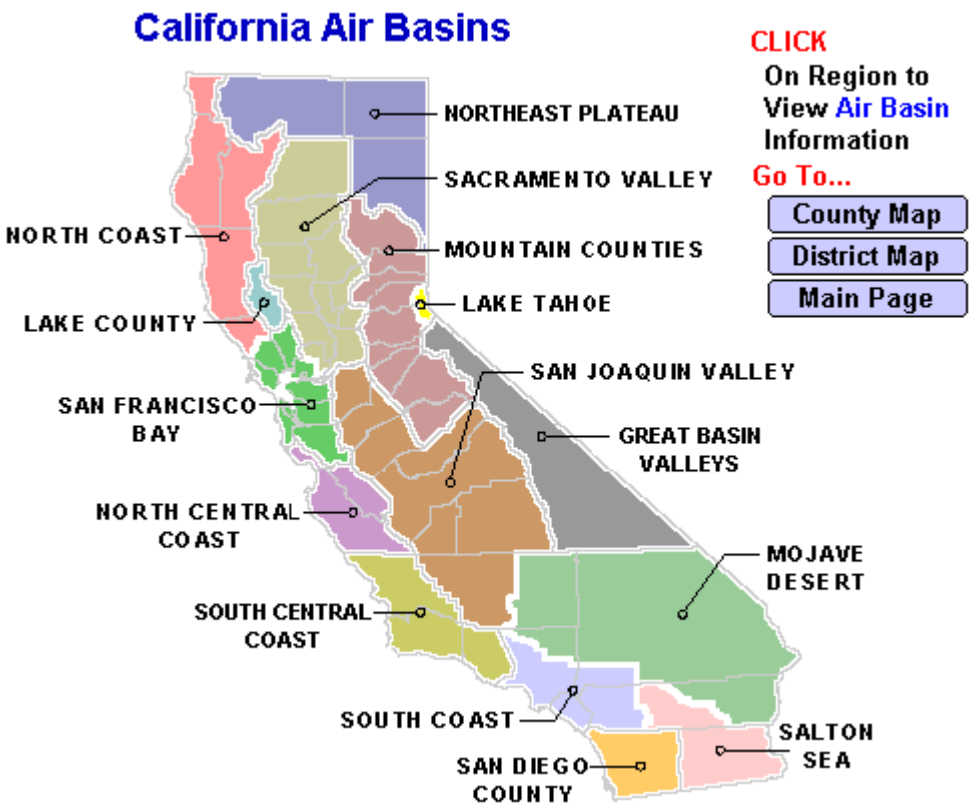
Zhang, M., Uno, I., Zhang, R., Han, Z., Wang, Z. and Pu, Y. (2006). "Evaluation of the Models-3 Community Multi-scale Air Quality (CMAQ) modeling system with observations obtained during the TRACE-P experiment: Comparison of ozone and its related species." *Atmospheric Environment* 40(26): 4874.

Zhang, Y., Vijayaraghavan, K. and Seigneur, C. (2005). "Evaluation of three probing techniques in a three-dimensional air quality model." *J. Geophys. Res.* 110: D02305, doi:10.1029/2004JD005248.

APPENDIX A:
Quality Assurance of CCOS Emissions

Emissions for the CCOS domain simulations were obtained from the California Air Resources Board (ARB) in five source categories: point, mobile, area, fire, and biogenic. These emissions were first summed and compared to the ARB emission inventory posted on its website for year 2000 (http://www.arb.ca.gov/ei/ei.htm).

Figure A1: California Air Basins



Source: ARB, http://www.arb.ca.gov/ei/maps/statemap/abmap.htm

The following air basins are included in calculating total emissions in ARB's inventory for the CCOS domain: North Coast, Lake County, San Francisco Bay, North Central Coast, South Central Coast, San Joaquin Valley, Great Basin Valleys, Mojave Desert, Mountain Counties, Lake Tahoe, and Sacramento Valley.

As ARB is using different source categories in their inventory, the following mapping is used for the purpose of comparison:

CCOS Emission Source Category	ARB Inventory (2000 Estimates from ARB Website)
Motor vehicle	On-road motor vehicles
Area	Total area-wide sources Other mobile sources (off-road source, ships, aircrafts, etc.)
Point	Stationary sources *
Biogenic	Biogenic sources (in Natural sources)
Fire	Wildfire (in Natural sources)

* Stationary sources in the ARB inventory include both point source and area source in the CCOS emission file, and therefore, to make a fair comparison, the sum of point and area sources should be compared between the ARB inventory and CCOS emissions.

Species Mapping

In ARB's inventory, emissions in tons/day are estimated for reactive organic gases (ROG), CO, NOx, SOx, etc. The CCOS emissions are speciated, and have units in moles/s. The following mapping is used in comparison. Note ARB's convention is to assume that all NO emissions are oxidized to NO_2. Therefore, they (and likewise the District) apply a molecular weight of 46 g/mol to NO. Therefore, I also applied 46 g/mol for all the NOx species.

Table A1: Mapping of Chemical Species

ARB Species	CMAQ Emissions	Molecular Weight (g/mol)
CO	CO	28
NOx	NO	46 (30)
	NO_2	46
	HONO	46 (47)
SOx	SO_2	64
	SULF	96
ROG	ETHENE	28
	ISOPRENE	68
	ALK1	30
	ALK2	35.6
	ALK3	45.4
	ALK4	76.2
	ALK5	102.8

ARB Species	CMAQ Emissions	Molecular Weight (g/mol)
	ARO1	94.4
	ARO2	114.4
	OLE1	44.8
	OLE2	60.2
	TRP1	136

Molecular weights of lumped VOC species are calculated based on estimated carbon number.

Table A2: Carbon Numbers for VOC Species

ETHENE 2
ISOPRENE 5 (biogenic)
ALK1 2.0
ALK2 2.4
ALK3 3.1
ALK4 5.3
ALK5 7.2
OLE1 3.2 (mostly propene; other C4+ terminal alkenes)
OLE2 4.3 (cis and trans-2-butenes, butadiene, and some C5+ internal alkenes)
ARO1 7.2 (mostly toluene, some ethylbenzene, not much else)
ARO2 8.7 (mostly C8-C10 aromatics with more than one alkyl group)
TRP1 10. (biogenic terpenes)

Comparison Results

Day-specific CCOS total emissions (weekday and weekend, respectively) and ARB's inventory are listed in the following figures in tons/day.

Figure A2: CCOS Emission Inputs Compared to ARB Inventory: Mobile Sources

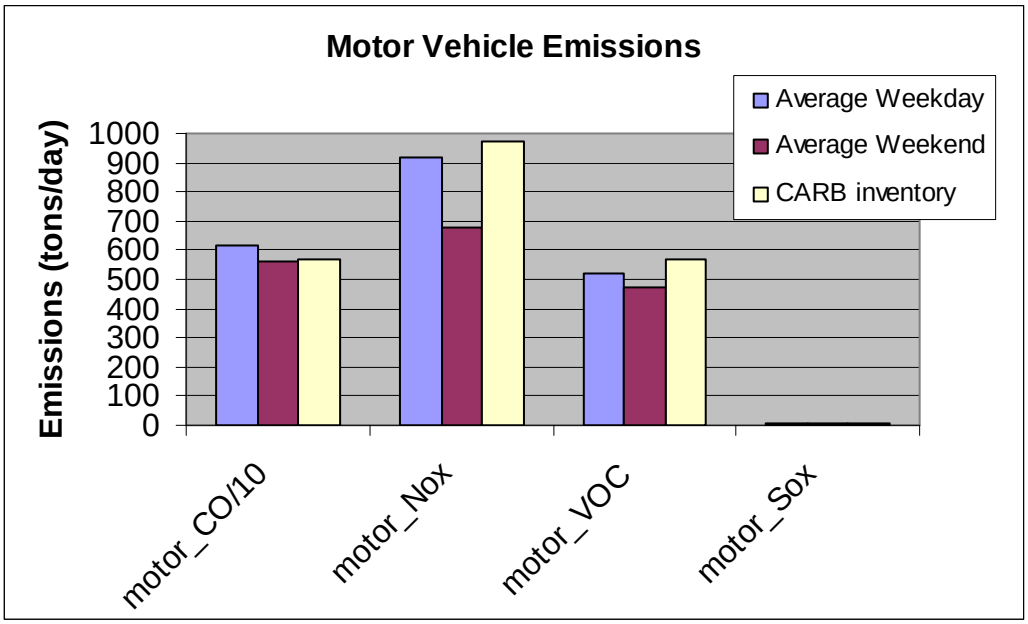


Figure A3: CCOS Emission Inputs Compared to ARB Inventory: Stationary Sources (Area+Point)

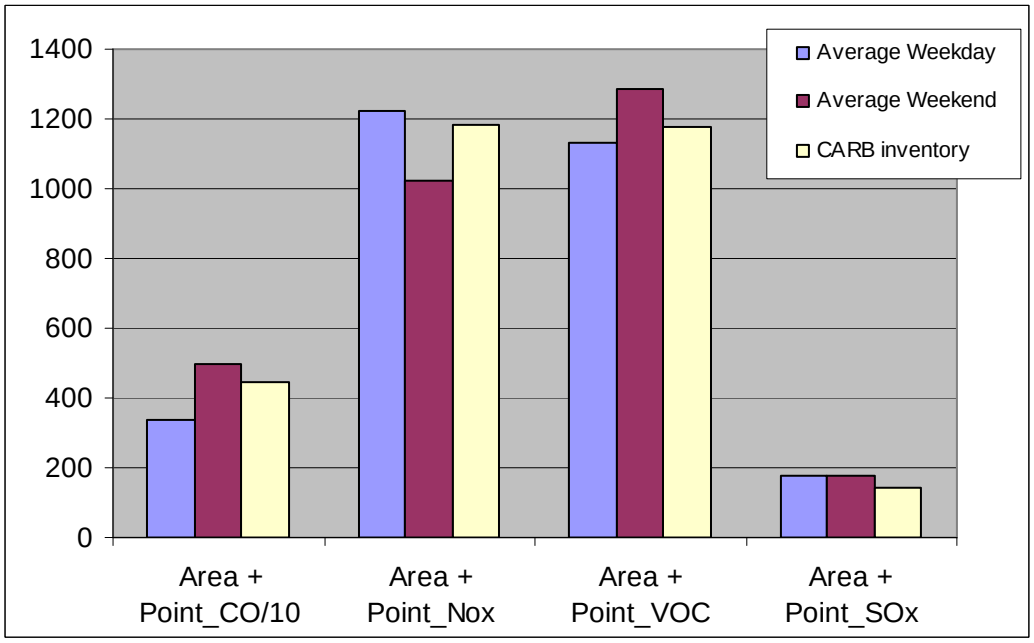
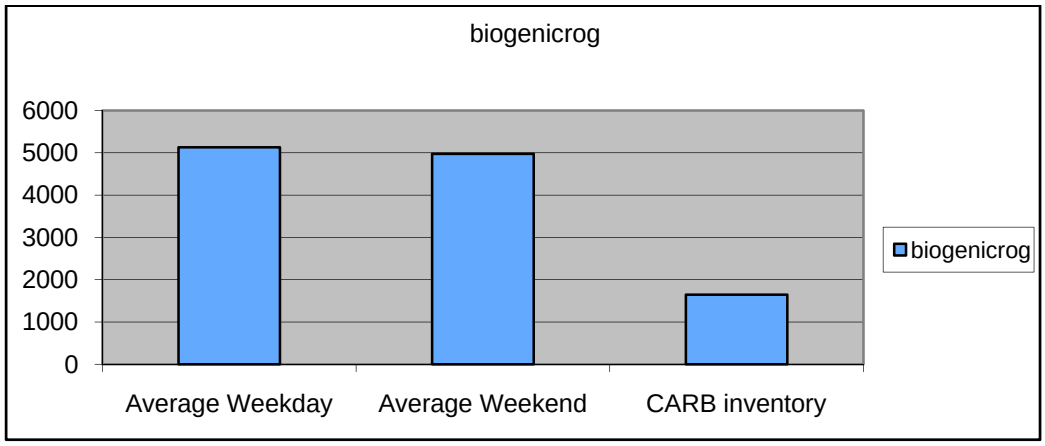


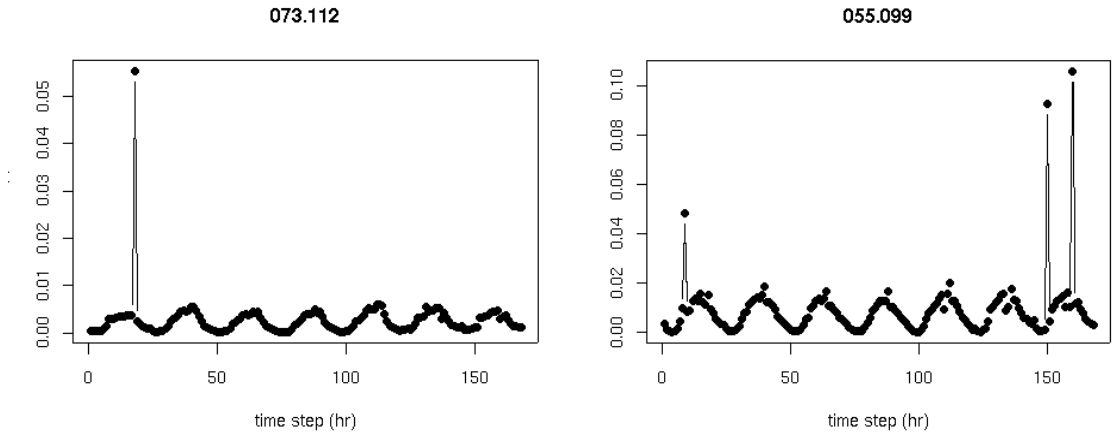
Figure A4: CCOS Emission Inputs Compared to ARB Inventory: Biogenic Sources (Tons/Day)



The magnitudes of all the anthropogenic emissions agree well between the two data sources. CCOS biogenic emissions are for the summer time, when higher temperature leads to greater emissions. It is hence reasonable to expect higher CCOS biogenic VOC emissions than ARB's annual average inventory (as seen in Figure A4).

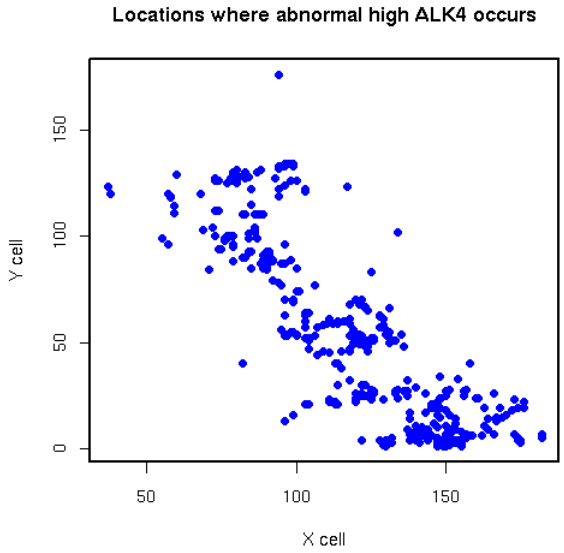
Six species exhibited abnormal temporal patterns in emissions (mol/s). Examples of abnormal temporal patterns (mol/s) in ALK4 emissions, selected by the criterion (max>5*mean). The figure title is in the format X.Y, where X and Y are the grid cell numbers.

Figure A5: Abnormal Temporal Profile in ALK4 from Mobile Sources



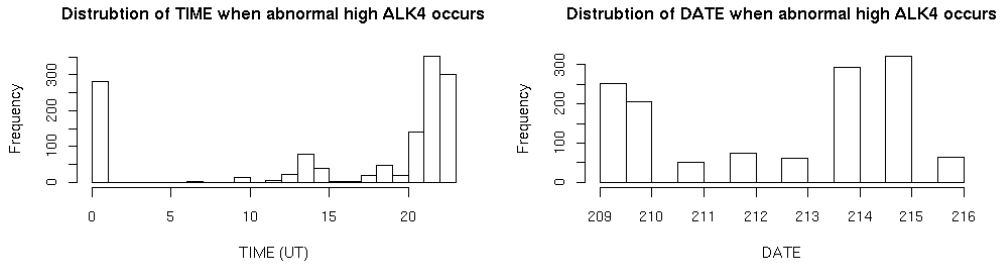
These patterns occur at number of locations.

Figure A6: Locations Where Abnormal Temporal Patterns Occur in ALK4 Emissions



The following two figures showed the distribution of time and date, when the spikes appear.

Figure A7: Distribution of Time and Date When the Abnormal Pattern Occurs



Note that the spikes mostly occur on the afternoon of weekdays (209, 213, 214), but not on all the weekdays. For seasonal modeling, we are using weekdays (Mon~Thu) all the same, so we selected a good weekday (i.e., with a minimum spike) and used its emissions for all other typical weekdays (Mon~Thu) in the summer, thereby eliminating the problem.

Double-counted fire emissions were found in two counties (Tuolumne and northern Fresno) in the area sources and these days. These days were replaced with other correct weekday emissions to remove the double-counted fire.

APPENDIX B:
Supporting Materials for Ozone Cluster Analysis

Here, a brief description of the procedure for K-means clustering analysis is provided.

First, find the initial partition of the ozone days and calculate the mean ozone vector for each cluster:

$$\mu_r = \frac{1}{N_r} \sum_{i \in r} \vec{X}_i \qquad \text{(B1)}$$

where *Nr* is the number of days in cluster *r*. Second, calculate the Euclidean distance between each day and each cluster mean. Reassign the day to the cluster that its mean is closest to. Repeat the first step and continue and the iteration until total within-group sum of squares (WSS) converges.

$$WSS_k = \sum_{r=1}^{k} \sum_{i \in r} (\vec{X}_i - \mu_r)^2 \qquad \text{(B2)}$$

The process is illustrated schematically in Figure B1.

Figure B1: Flow Diagram Describing the Clustering Procedure

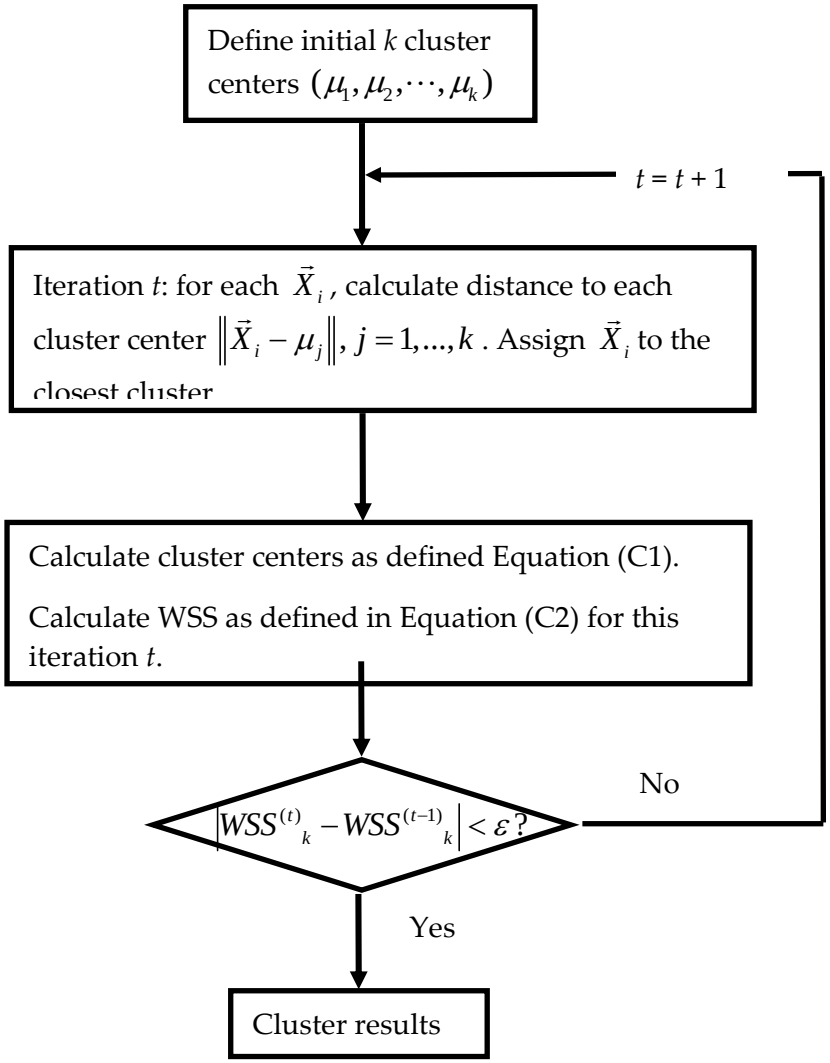


Figure B2: Variance and Cumulative Variance Explained by Principal Components of 8-h Ozone Maxima in Each Subregion: (a) SJV, (b) SFB, (c) SV

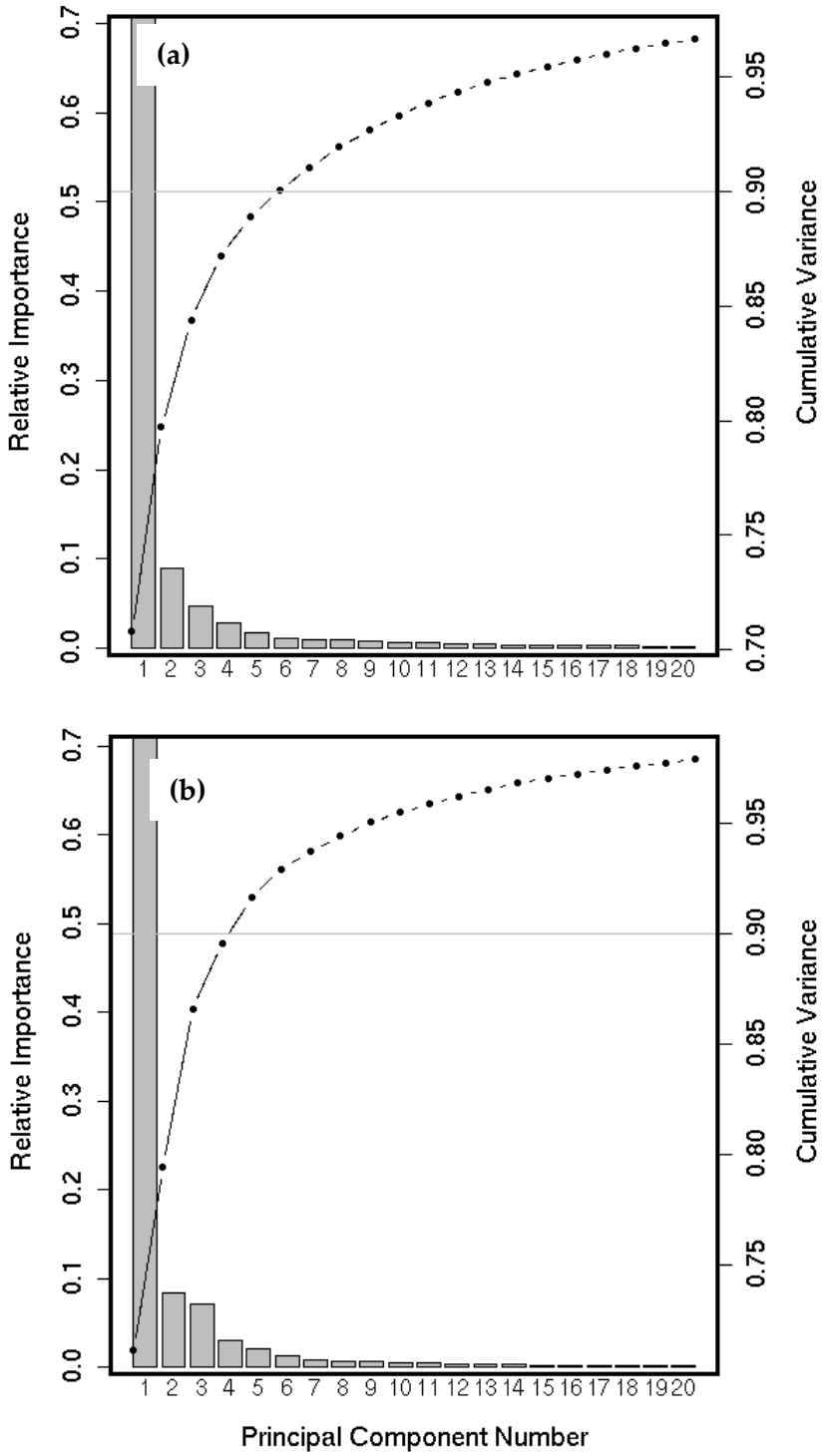
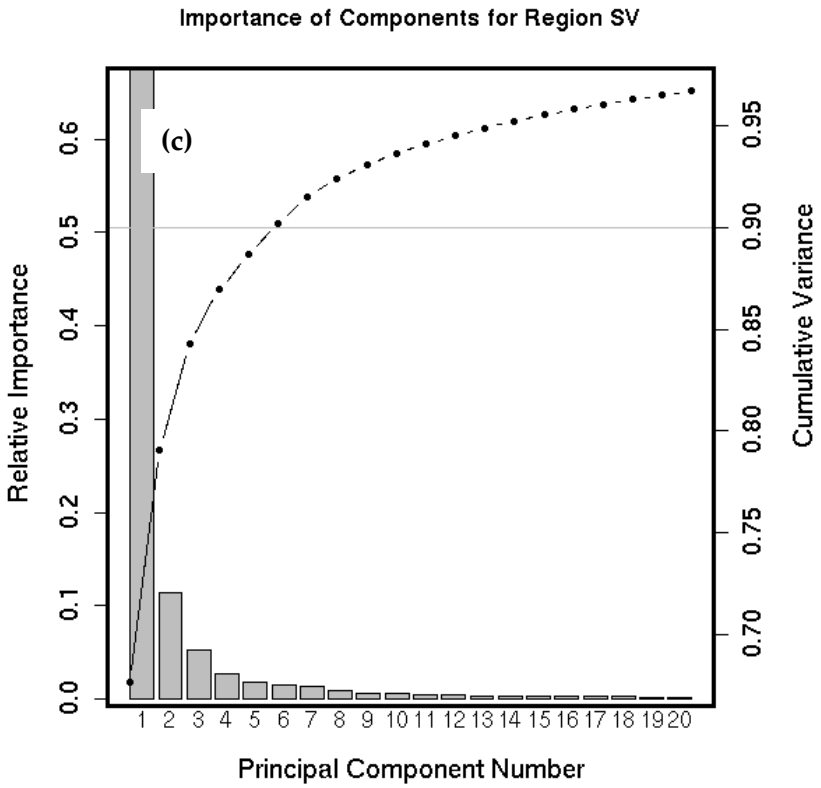


Figure B2: Continued



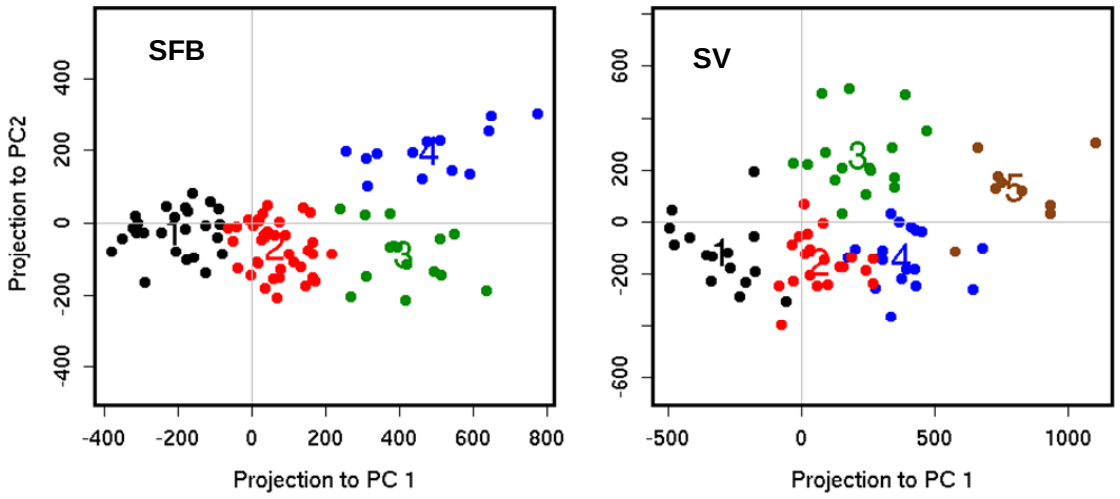
To view the ozone clusters in a more dynamic way, a transition matrix is presented in Table B1, showing the number of times each cluster made a transition either to itself or to another clusters. Clean days (cluster 0) are also presented here. The last column in Table B1 is the persistence ratio, the number of times a cluster followed itself to the number of times the cluster made a transition to a different cluster. The larger the persistence ratio, the more stable the clusters. The largest two numbers in each row are colored in red, indicating the two most probable transitions. Low-ozone days (clusters 0, 1, and 2) are less likely to develop into high-ozone clusters (4, 5, and 6) on the next day, which implies that the ozone accumulation process in the SJV usually takes multiple days. Among the clusters that have relatively high ozone (clusters 3-6), clusters 4 and 5 seem to be more persistent than 3 and 6, which may indicate that the meteorological regime associated with these clusters tends to persist for more than one day before transitioning to another cluster. In contrast, clusters 3 and 6 are shorter-lived. Cluster 3 is mostly developed from cluster 2 and transitions to cluster 5. The transition matrix indicate the two most probable paths for SJV ozone levels to develop from low to high: 1→4→6 (then fall back to 5), or 2→3→5 (then fall back to 2, or, less likely, develop into 6). As seen in Figures 5-4 and 5-5, clusters 2 and 5 have similar spatial contrast features, as characterized by PC2, but different average ozone levels as characterized by PC1. Similarly, clusters 1 and 4 have similar spatial contrasts but different ozone levels. The two paths (1→4→6, and 2→3→5) may indicate

the development of two different weather regimes. Statistical significance of these speculations should be further investigated with larger sample sizes.

Table B1: Transition Matrix for SJV Ozone Clusters

	To Cluster							
From Cluster	0	1	2	3	4	5	6	Persistent Ratio
0	20	3	4	1	0	0	0	2.5
1	1	4	1	0	2	0	0	1.0
2	3	0	8	7	1	3	0	0.6
3	2	0	3	4	2	4	1	0.3
4	0	1	0	1	7	1	3	1.1
5	1	1	5	2	0	8	3	0.7
6	0	0	1	1	1	4	3	0.4

Figure B3: Scatter Plots of Ozone Days on PC1 and PC2. Cluster indices are shown in the same color as the associated days at the respective cluster centers.



SFB clusters 1, 2, 3 have similar PC2 scores and are separated by the mean ozone anomaly levels indicated by PC1 scores. SFB cluster 4 is different from the others by spatial contrast (higher PC2 scores). SV clusters 1, 2, 4 are separated by the mean ozone anomaly levels indicated by

PC1 scores. SV cluster 5 has the highest ozone levels, as well as different spatial contrast from 1, 2, and 4. SV cluster 3 differs from the others by higher scores on PC2.

Figure B4: Average Daily Maximum Temperature (K) for Summer 2000 and the Temperature Loadings (K) on the First Three PCs

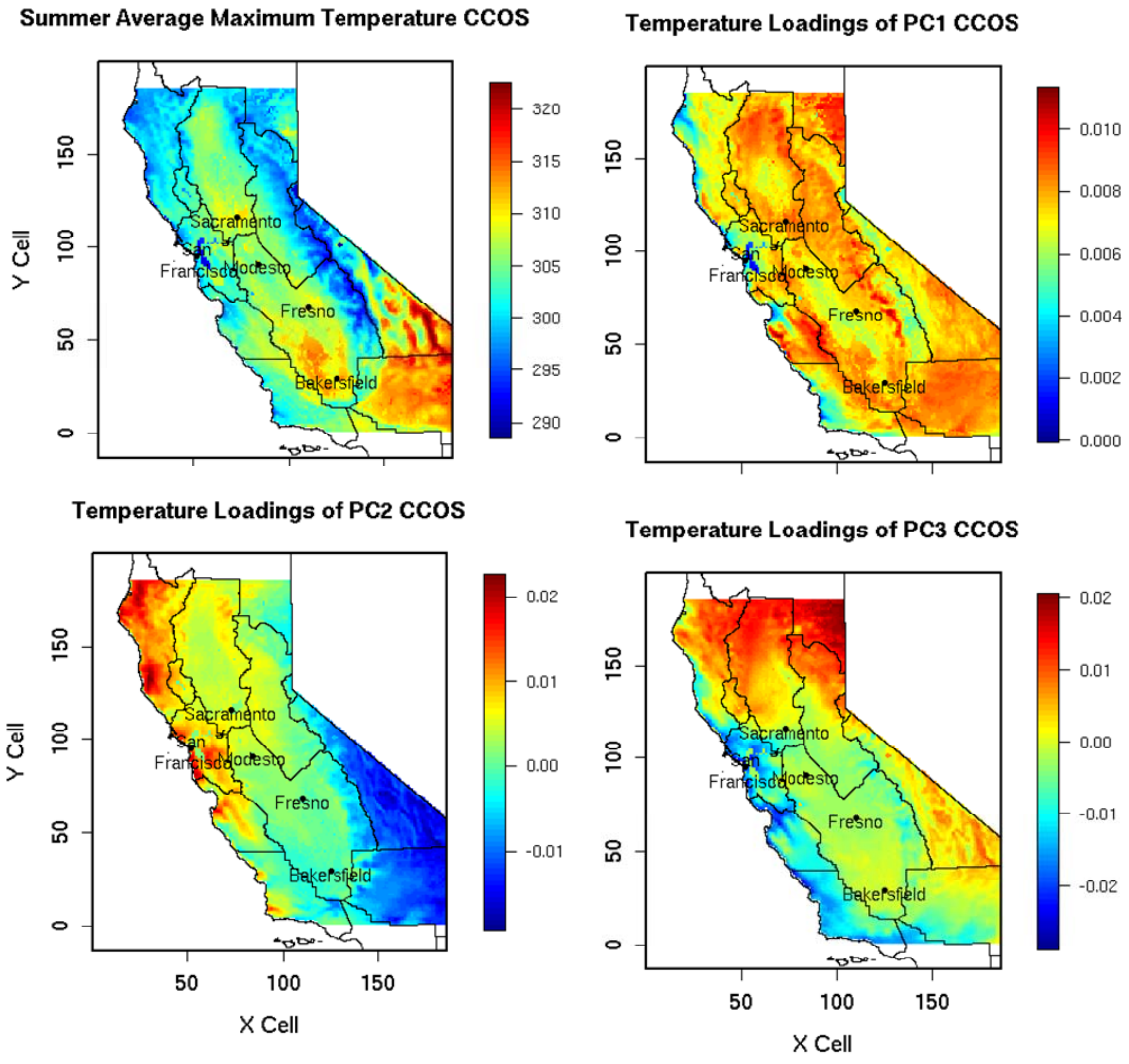


Figure B5: Variances and Cumulative Variances Contributed by Wind PCs

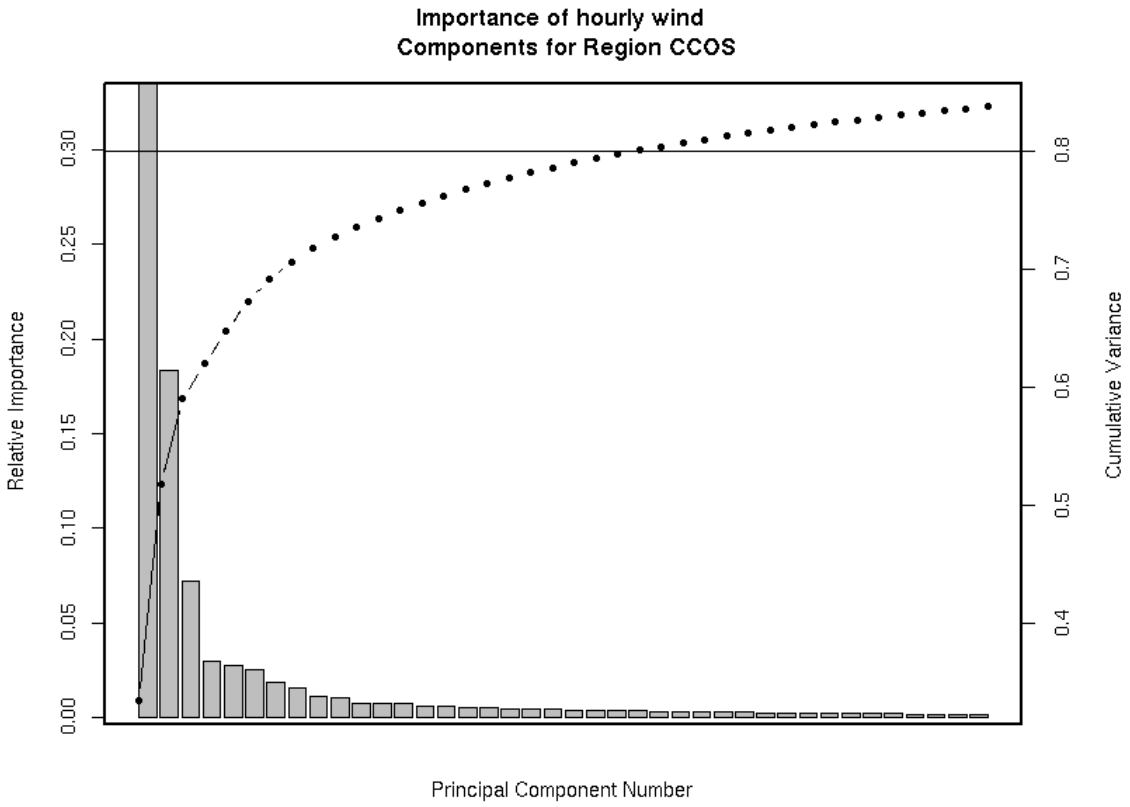
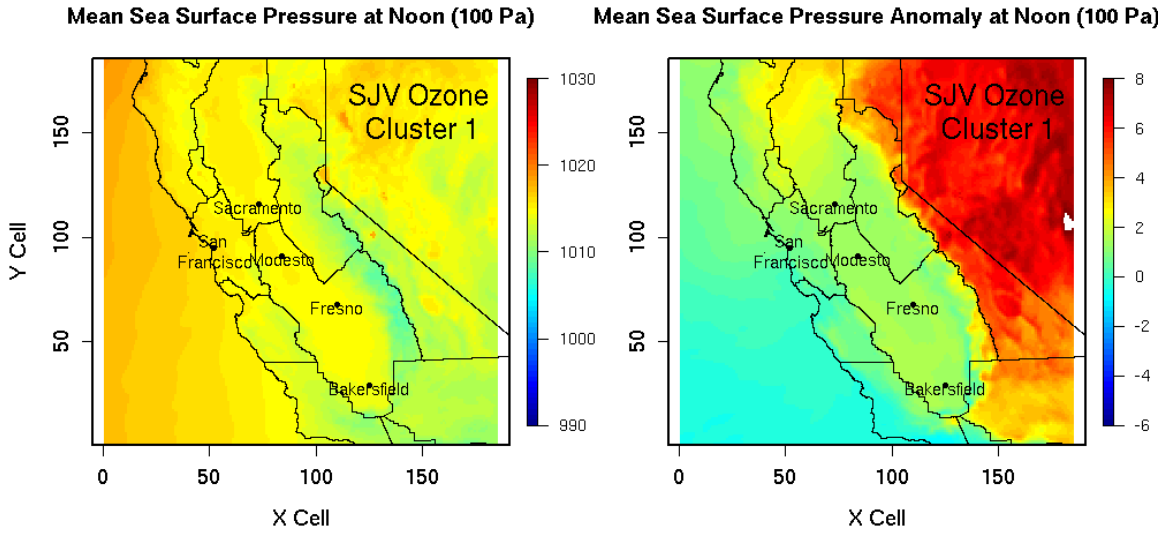
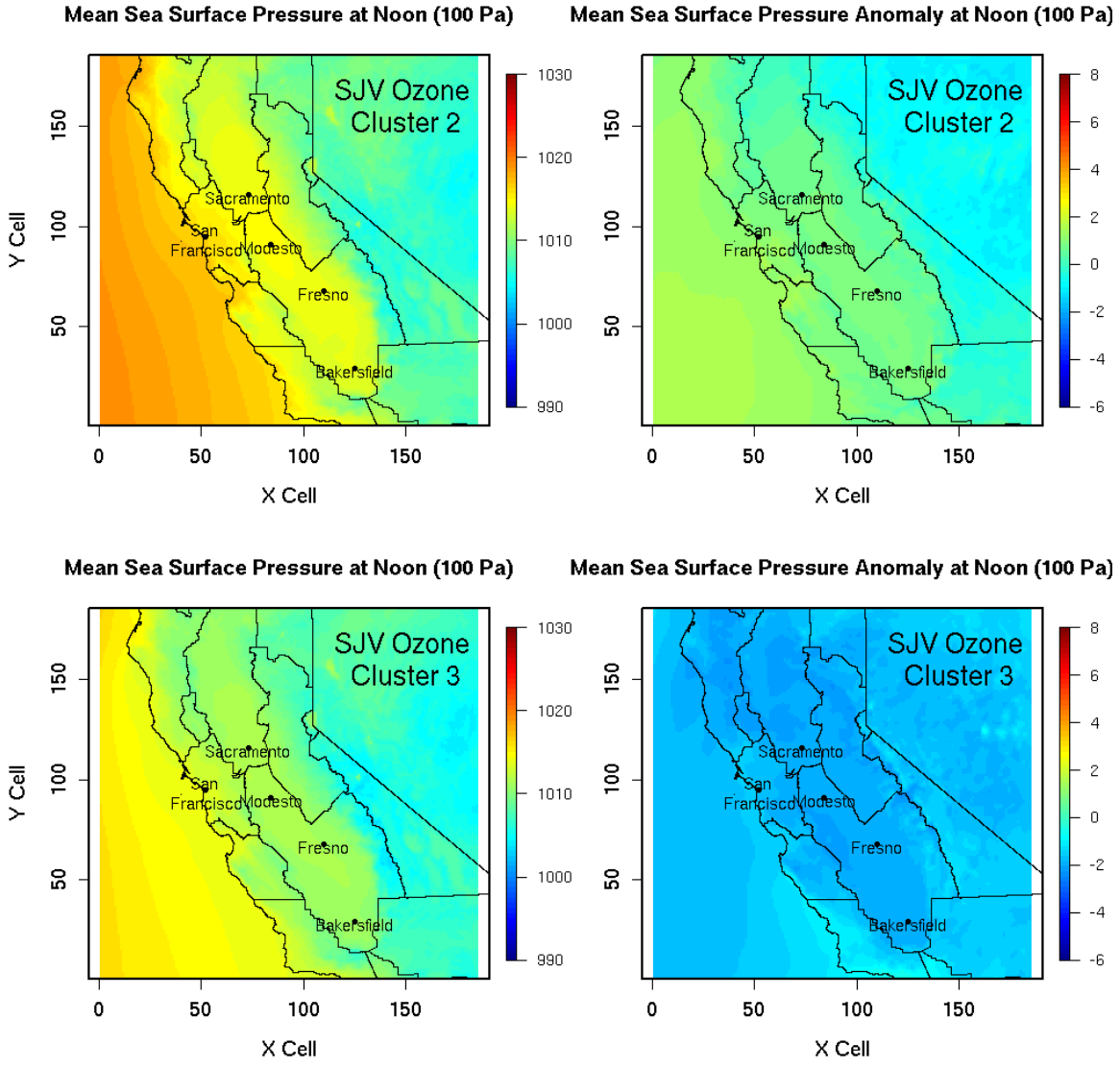
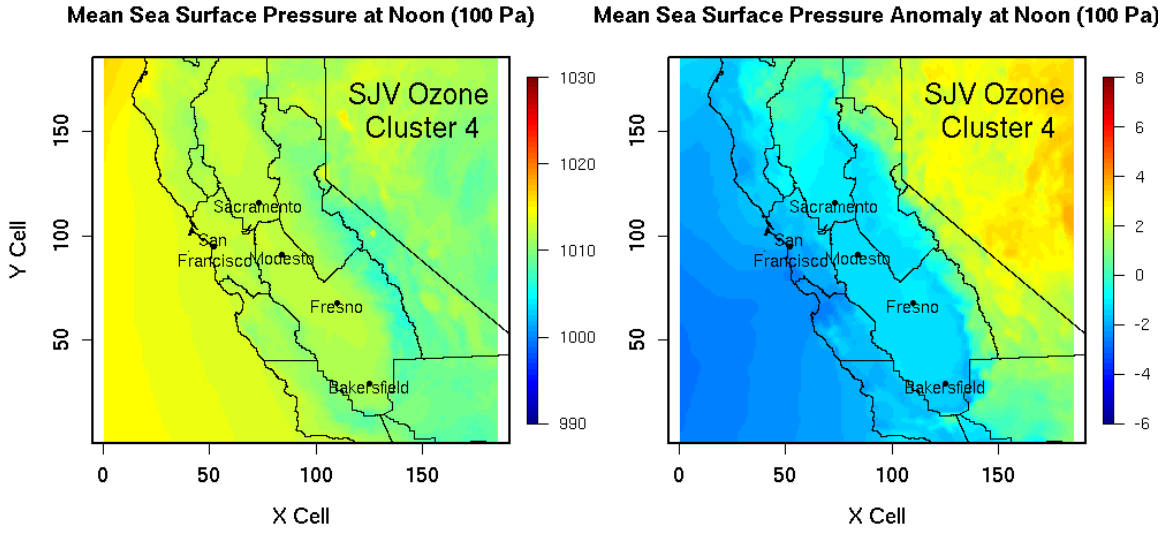


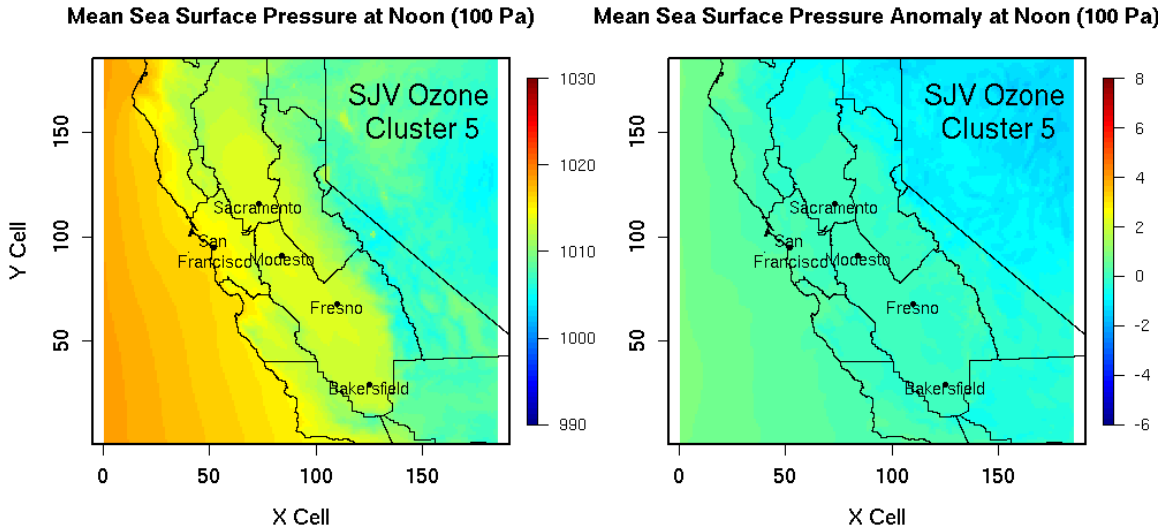
Figure B6: Mean Sea Surface Pressures and Anomolies (at Noon) Averaged Over Each SJV Cluster




Mean Sea Surface Pressure at Noon (100 Pa)
Mean Sea Surface Pressure Anomaly at Noon (100 Pa)
SJV Ozone Cluster 2
SJV Ozone Cluster 3
Sacramento
San Francisco
Modesto
Fresno
Bakersfield
Y Cell
X Cell
0
50
100
150
990
1000
1010
1020
1030
-6
-4
-2
0
2
4
6
8


Mean Sea Surface Pressure at Noon (100 Pa)
SJV Ozone Cluster 4
Sacramento
San Francisco
Modesto
Fresno
Bakersfield
Y Cell
X Cell
0
50
100
150
990
1000
1010
1020
1030

Mean Sea Surface Pressure Anomaly at Noon (100 Pa)
SJV Ozone Cluster 4
Sacramento
San Francisco
Modesto
Fresno
Bakersfield
X Cell
0
50
100
150
-6
-4
-2
0
2
4
6
8


Mean Sea Surface Pressure at Noon (100 Pa)
SJV Ozone Cluster 5
Sacramento
San Francisco
Modesto
Fresno
Bakersfield
Y Cell
X Cell
0
50
100
150
990
1000
1010
1020
1030

Mean Sea Surface Pressure Anomaly at Noon (100 Pa)
SJV Ozone Cluster 5
Sacramento
San Francisco
Modesto
Fresno
Bakersfield
X Cell
0
50
100
150
-6
-4
-2
0
2
4
6
8